Special Proceedings of the Korean Scientists and Engineers Association in Germany

Vol.1

재독한국과학기술자협회
Verein Koreanischer Naturwissenschaftler und Ingenieure in Deutschland
Korean Scientists and Engineers Association in Germany

Editorial Board

Dongwoon Bae

ENP International
Auf dem Alten Kamp 20
44803 Bochum

Chan Kim, Ph.D.

European XFEL GmbH
Holzkoppel 4
22869 Schenefeld

Jitae Park, Ph.D.

Heinz Maier-Leibnitz Zentrum (MLZ)
Technische Universität München
Lichtenbergstr 1
85748 Garching bei München

Hyun-Woo Jeong, Ph.D.

Single Cell Multi-Omics Laboratory
Max-Planck Institute for Molecular Biomedicine
Röntgenstraße 20
48149 Münster

Sungah Kang, Ph.D.

Institute of Climate and Energy Systems (ICE)
Troposphere (ICE-3)
Wilhelm-Johnen-Straße
52425 Jülich

Min Son, Ph.D.

Institut für Thermodynamik
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39
85577 Neubiberg

Junbeom Park, Ph.D.

Institute of Energy Technologies - Fundamental
Electrochemistry (IET-1)
Forschungszentrum Jülich
Wilhelm-Johnen-Straße
52425 Jülich

Dr. -Ing. Jeongwook Koh

Renesas Electronics Germany GmbH
Grenzstraße 28
01109 Dresden

Kwangjin Lee, M.Sc. Architekt

K LEE ARCHITEKT
Sachtlebenstrasse 53
14165 Berlin

Yerin Baek

Leibniz Universität Hannover
Welfengarten 1
30167 Hannover

Individual contributions:
Copyright © 2026 by the respective authors. All rights reserved.

Compilation and editorial design:
Copyright © 2026 VeKNI. All rights reserved.

특별 학술 논문집 발간사

재독한국과학기술자협회 회원 여러분, 그리고 내외 귀빈 여러분, 협회의 자랑스러운 유산이 될 이번 특별 학술논문집 발간을 진심으로 축하드립니다.

특히 지난 6 년간 재독과협회장으로서 협회를 이끌어온 저에게 이번 특별 학술논문집 발간은 개인적으로도 매우 남다르고 뜻깊은 이정표입니다. 지난 6 년의 임기 동안 우리 협회는 EKC(유럽학술대회), 전문분과 활동, 그리고 각 지역회 활동 등을 통해 그 어느 때보다 활발하게 학술 교류를 이어왔습니다. 이번 특별 논문집은 그동안 우리가 함께 일구어온 수많은 학술적 성과와 열정을 하나로 집대성한 값진 결실입니다.

과학과 기술은 언제나 호기심과 협업, 그리고 국경을 초월한 지속적인 아이디어 교류를 통해 발전해 왔습니다. 세계 최고의 연구 및 혁신 중심지인 이곳 독일은 우리 한국인 과학기술자들이 우수한 연구를 수행하는 동시에, 전 세계 동료들과 견고한 파트너십을 구축할 수 있는 훌륭한 토대가 되어주었습니다. 우리 회원 여러분이 이뤄낸 성과는 인류 과학의 발전에 기여할 뿐만 아니라, 대한민국과 독일 간의 오랜 우호 관계를 더욱 공고히 하는 가교 역할을 하고 있습니다.

이번 특별 논문집은 재독 한국인 과학계의 다양성과 탁월함을 여실히 증명하고 있습니다. 본 논문집에 수록된 연구들은 폭넓은 학문적 스펙트럼을 아우르고 있으며, 우리 회원들의 높은 연구 수준과 과학기술의 지평을 넓히기 위한 끊임없는 헌신을 잘 보여줍니다. 동시에 이 간행물은 재독 한국인 연구자들이 펼쳐온 활발한 연구 활동의 소중한 기록이자, 국제 과학 협력의 중심에서 우리가 수행하고 있는 핵심적인 역할을 보여주는 증거이기도 합니다.

저는 이 논문집이 연구자들 간의 소통을 더욱 긴밀히 하고, 새로운 융합 협력을 촉진하며, 해외에서 연구 커리어를 쌓고자 하는 차세대 젊은 과학자들에게 영감을 주는 든든한 플랫폼이 되기를 기대합니다. 오늘날 우리는 기후변화로 인해 전 인류가 위험에 노출되고 있는 현장을 목격하고 있습니다. 우리 사회가 직면한 거대한 문제들로 인해 인류는 큰 위기에 처해 있으며, 이는 결코 한 개인이나 국가의 힘만으로는 해결할 수 없습니다. 이 거대한 과제들은 오직 글로벌 협력을 통해서만 돌파구를 찾을 수 있습니다. 그러하기에 이번 논문집 발간은 우리에게 강력한 국제 과학 네트워크를 유지하고 발전시키는 것이 얼마나 중요한지 다시 한번 일깨워줄 것입니다.

이 특별한 논문집이 세상에 나오기까지 시간과 노력을 아끼지 않으신 김찬 학술간사님과 각 분과장님들을 비롯한 모든 관계자분들께 고개 숙여 깊은 감사의 말씀을 전합니다. 아울러 바쁜 연구 일정 속에서도 흔쾌히 소중한 연구 논문을 투고해 주신 회원 여러분께도 진심으로 감사드립니다. 여러분의 공동 노력 덕분에 이 특별한 논문집이 결실을 맺었을 뿐만 아니라, 재독 한국인 과학계의 우수성, 다양성, 그리고 협동정신을 담은 위대한 기록을 남길 수 있었습니다. 이 논문집은 여러분의 업적을 기리는 영원한 증표가 될 것이며, 학문과 국경을 초월한 과학적 교류를 계속해서 고취할 것이라 확신합니다

이번 특별 학술논문집 발간을 도약대 삼아 우리 협회가 양적·질적으로 한 단계 더 성장하기를 바라며, 앞으로도 재독 한국인 과학계의 결속을 다지고 한·독 간의 과학기술 협력을 더욱 증진하는 데 지속적으로 기여할 수 있기를 뜻깊은 마음으로 기원합니다. 감사합니다.

재독한국과학기술자협회 31, 32 대 회장

배동운

Foreword

It is with great pleasure that we present these inaugural Special Proceedings of the Korean Scientists and Engineers Association in Germany (VeKNI). This publication is the result of the dedication, enthusiasm, and collective efforts of many researchers who generously contributed their work to celebrate the vibrant Korean scientific community in Germany.

Science and engineering continue to evolve through curiosity, creativity, and collaboration. In today's rapidly changing research environment, meaningful advances are increasingly achieved by bringing together diverse expertise across disciplines, institutions, and national boundaries. As member of an international scientific community, Korean scientists and engineers in Germany have actively contributed to these developments while strengthening academic and technological cooperation between Korea, Germany, and the global research community.

These proceedings reflect the breadth and diversity of research conducted by our members. The articles presented here span a wide range of disciplines, including physics, chemistry, biology, materials science, engineering, medicine, computer science, artificial intelligence, and many others. Although these fields may appear distinct, they are united by a common goal: advancing scientific knowledge and developing innovative solutions that contribute to society.

Beyond introducing recent research achievements, we hope this publication serves as a platform for communication and collaboration among researchers. We also hope it encourages interdisciplinary discussions, inspires early career scientists, and strengthens the network of Korea researchers both within Germany and around the world. We believe that sharing scientific ideas and experiences is just as valuable as sharing research results, and that a strong research community is built through openness, mutual respect, and lasting partnerships.

These special proceedings would not have been possible without the dedication of many individuals. We extend our sincere gratitude to all contributing authors for sharing their outstanding work and research experiences, to the reviewers for their thoughtful comments, and to the editorial committee for their unwavering commitment throughout the publication process. Their efforts have made these proceedings a meaningful record of the achievements and collaborative spirit of the Korean scientific community in Germany.

We sincerely hope that these special academic proceedings will serve not only as a valuable scientific resource but also as a lasting symbol of friendship, collaboration, and shared aspirations among Korean scientists and engineers. May it inspire new ideas, new partnerships, and new discoveries in the years ahead.

Editorial Committee

Special Proceedings of the Korean Scientists and Engineers Association in Germany (VeKNI), Vol. **1** (2026)

Celebrating Scientific Excellence and Collaboration

It is a great pleasure to congratulate the Korean Scientists and Engineers Association in Germany (VeKNI) on the publication of this special academic volume.

Since its launch in 1973 as the very first Korean scientific association in Europe, VeKNI has served as a vital platform for meaningful connection and knowledge sharing for over half a century. This special volume is a truly meaningful achievement, designed to share the outstanding research accomplishments of VeKNI members and to provide a useful guide for the next generation of scientific talents. Furthermore, this publication arrives at a particularly welcome and significant moment as we celebrate the 40th anniversary of the Korea-Germany Scientific and Technological Cooperation Agreement this year.

The articles collected here represent a broad spectrum of research fields, demonstrating both the high quality of their work and their dedication to advance science and technology. At the same time, this publication serves as a valuable record of the vibrant research activities carried out by Korean researchers and highlights the important role they play in international scientific collaboration. I sincerely hope that this volume will become more than a collection of scientific articles. It should also serve as a platform that encourages communication among researchers, promotes new collaborations, and inspires young scientists who aspire to pursue research careers abroad.

This special volume would not have been possible without the dedication and commitment of many individuals. I would like to extend my deepest appreciation to everyone whose time and effort made this publication a reality. Their collective commitment has not only brought this special volume to fruition but has also created a meaningful record of the excellence, diversity, and collaborative spirit of the Korean scientific community in Germany. I am confident that this volume will serve as a lasting testament to their achievements and continue to inspire scientific exchange and cooperation across disciplines and borders.

May this special academic volume strengthen the bonds within the Korean scientific community in Germany while further promoting scientific cooperation between Korea and Germany for many years to come.

Chul-hwan Byun

Head of the Embassy of the Republic of Korea, Bonn Office

Connecting Ideas, Inspiring Innovation

It is a great honor and pleasure to congratulate the Korean Scientists and Engineers Association in Germany (VeKNI) on the publication of this special academic volume. This publication stands as a true testament to the dedication, excellence, and collaborative spirit of Korean researchers in Germany.

Scientific progress has never been confined to a single discipline. Many of today's most exciting discoveries emerge from the convergence of diverse fields, where researchers with different expertise come together to exchange ideas, challenge conventional thinking, and create new opportunities for innovation. Equally important is the international nature of modern research, which allows knowledge to move freely across institutions, cultures, and national borders.

The Korean scientific community in Germany represents this spirit exceptionally well. Researchers from a wide range of disciplines—including physics, chemistry, biology, engineering, medicine, materials science, artificial intelligence, and many other fields—have established themselves as active members of the global research community. While pursuing excellence in their respective areas, they also serve as important bridges connecting Korea and Germany through scientific collaboration, academic exchange, and human networks.

This special volume beautifully reflects the diversity of these research activities. More than simply presenting scientific achievements, it captures the collective efforts of Korean researchers who continue to contribute to the advancement of science and technology in many different ways. It also demonstrates that meaningful innovation often begins with communication—sharing ideas, learning from one another, and building collaborations that extend beyond individual disciplines.

It is my sincere hope that this publication serves as a valuable platform not only for introducing outstanding research results but also for strengthening connections within the Korean scientific community in Germany. Furthermore, I hope it fosters new interdisciplinary collaborations, inspire early-career researchers, and broaden horizons for international cooperation.

I would like to express my sincere appreciation to the editorial committee, the contributing authors, and everyone who devoted their time and effort to making this publication possible. Their dedication has created a lasting record of the vibrant Korean research community in Germany and its continuing contributions to global science and technology. I sincerely hope that this special volume continues to nurture collaboration, and friendship among researchers, while further strengthening the scientific partnership between Korea and Germany for the benefit of future generations of researchers.

I also wish to extend my deepest gratitude to the Korean scientists and engineers in Germany. Despite the challenges of working far from their home country, their unwavering dedication has played a pivotal role in propelling Korea's scientific and technological advancement. I sincerely wish them continued growth and even greater success in all their future endeavors.

Do Young Noh

Professor of Physics and Photon Science, Gwangju Institute of Science and Technology (GIST)

Former President of the Institute for Basic Science (IBS)

Contents

Part 1: Physics Division

Monolithic III-V Nanowire Quantum Dot Emitters for Integrated Silicon Photonics 2
Hyowon Jeong and Gregor Koblmüller

Electrical polarization and probe characteristic of ferroelectric van der Waals material α-In2Se3 7
Sunwoong Yang, Minwoo Lee, Young-Mi Bahk and Sung Ju Hong

X-ray Photon Correlation Spectroscopy: Probing Nanoscale Dynamics with Coherent X-rays 11
Wonhyuk Jo

Development of periodic Hirshfeld atom refinement 15
Kanghyun Chu

Recent Advances in Single-Particle Imaging at XFEL Facilities 19
Chan Kim

Environmental Liquid Cell for Three-Dimensional Nanoscale Control of Liquid Water 23
Chiwon Lee

Magneto plasmonic responses in periodic structure 27
Jeong Woo Han

Hydrogen Lyman-α Emission in Astrophysics 30
Seok-Jun Chang

Part 2: Life Science Division

Visualizing Proteins at Work: Time-Resolved Serial Crystallography and Sample Support at the European XFEL 36
Huijong Han

Convergence of Single-Cell Omics and Spatial Omics 40
Hyun-Woo Jeong

Fates and Generation of Supernumerary Centrioles 43
Byungho Shin

Beyond Conventional Fluorescence: Bringing Quantitative Microscopy into Routine Biological Research 47
Kyoohyun Kim

From Cell Fate Curiosity to Everyday Genomics 51
Rinho Kim

Technologies for Reproductive Capacity Regulation in Humans and Animals 55
Youn-Su Lee

Somatic Mosaicism Complicates Early Biomarker Validation: A Case for Longitudinal, Within-Person Monitoring 61
Hongryeol Park

Unresolved links between ammonia stress resilience and biomass production in microalgae grown in piggery wastewater 68
Eun Bin Joa and Sang-Ah Lee

Part 3: Environment and Energy Division

Maintenance Challenges and Inspection Technologies for Fixed-Bottom Offshore Wind Substructures in Europe 76
Martin S. D. Yoo and Jens Cruse

Highly Oxygenated Organic Molecules in Atmosphere: Formation, Detection, and Role in Secondary Organic Aerosol 81
Sungah Kang

Part 4: Mechanical Engineering Division

Learning from Nature: Biomimetic Applications of Owl-Inspired Serrations for Propeller Noise Reduction 86
Changjin Hong and Jong-Seob Han

Environmental Impact of Rocket Launches in New Space Era 89
Min Son

Engineering Additively Manufactured Injectors towards Fuel-Flexible and Scalable Combustion Technologies 92
Yeonse Kang and Fabian Hampp

Towards Learned Search Guidance for 3D Semantic Voxel Maps for Aerial Navigation 96
Taehyoung Kim

The Hidden Technology Behind Data Centers: Laser Processing 100
Woo-Sik Chung

Design of Stamping Tool Geometry for Metallic Bipolar Plates Using Bayesian Optimization 103
LANGENSTÜCK-LEE Sangwook, PARK Hyunjin, HÜSING Florian and BRECHER Christian

Federated Data Ecosystems as an Enabler of the Circular Economy- Manufacturing-X 108
Ro, Seonhi, Schmidt, Lara, Dietrich, Katrin, Revellio, Ferdinand, Mügge, Janine and Vollenbruch, Paul

A Proposal on the Disciplined Use of Large Language Model (LLM) AI for Engineering Research and Code Development 113
Junsu Shin

Part 5: Material Division

Mechanism-Guided Selectivity Control in Metal-Nitrenoid Transfer 118
Suhyeon Kim

Orthogonal B/Si Activation of -Silyl Allylboronic Esters for Stereodivergent Allyl–Allyl Coupling 122
Yongsuk Jung

4D-STEM Analysis of Process-Dependent Cu-Sn Intermetallic Evolution in Micro-Bumps for Advanced Semiconductor Packaging 125
Sangjun Kang

Advanced Synchrotron X-ray Scattering Facility and Operando Materials Dynamics at the PLS-II 3C SAXS-I Beamline 128
Young Yong Kim

Study of Electrochemistry via In-situ Transmission Electron Microscopy 132
Junbeom Park

Powering the world of microrobots with micro energy systems 135
Yeji Lee, Vineeth Kumar Bandari and Oliver G. Schmidt

Band-Structure Engineering of TiO2 with Pt Single Atoms for Photocatalytic Nitrate Reduction to Ammonia 138
Hayoon Jung

Accelerating Alloy Design using AI: It's Optimization All the Way Down 142
Jin-Young Lee

Part 6: Electrical, Electronics & Information Technology Division

A 22 nm FD-SOI CMOS Fully Differential Two Stage Folded-Mesh CLASS AB Operational Amplifier with an Internal Regulator 148
Jeongwook Koh

From Head Trajectory Diagrams to Deep Gait Recognition: A Single Head-Worn Sensor Approach to Gait Symmetry Visualization and Person Identification 155
Tong-Hun Hwang and Alfred O. Effenberg

From Sinusoids to Chirps: Rethinking the Wireless Physical Layer for 6G and Beyond 159
Hyeon Seok Rou

Unearthing Silent Data with Process Design 164
Jae Sook Cheong

Network-Enabled Distributed Physical AI: A 5G-Advanced and 6G Architecture for System 0–3 Robotic Intelligence 168
Harrison Hongseo Yun

Benchmarking Oversampling Techniques and Implementing a Binary Classifier for Identifying RuBisCO Protein Sequences 183
Ji Sun Kwon, Sonja Diedrich and Jochen Blom

Topological Features for Environmental Sound Classification: A Proof of Concept toward Industrial Acoustic Anomaly Detection 187
Jiwon Seo, Sascha Grollmisch, Jaehyun Lee, Sukjin Hong and Jooyoung Hahn

Enhanced Operational Characteristics of AlGaN/GaN HEMTs Using HfO_2-Based Passivation: A Simulation Study 191
Junho Lee and Hyun-Seok Kim

Part 7: Design and Architecture Division

Transformation of Berlin's Urban Structure through Long-Distance Railway Network Evolution: A Diachronic Analysis of Major Stations from a Spatial Metabolism Perspective 198
Sungdong Lee

Polycentricity: Historical Transformations and Urban Regeneration in Central Berlin 201
Kwangjin Lee

A Study on the Urban Spatial Transformation of Berlin's Embassy and Government Districts in Response to Political and Social Changes 205
Inho Woo

The Formation of Cologne’s Historic City Walls, the Green Belt, and the Contemporary Urban Structure 209
Soonkyo Jeong

Value Analysis in the World Heritage Convention: recognition and interpretation of the 'Criterion (ii)' on interchange of human value. 213
Semina An

MLK Models and Industrialized Architecture in the GDR 218
Sung-Yong Kim

The 15-Minute Play City: Infrastructure for Friction and Resonance in the Age of AI 223
Jaewook Kwon

Hazardous Material Management in German Building Renovation and Demolition Projects 227
Jin Hyun Park

Das Badhaus in Aachen 231
Minkyu Kim

Constructing a European Smart City: Digital Infrastructure, Urban Services, and Local Implementation in Berlin 235
Eunyoung Jang

Part 8: Experiences and Advice from Scientists and Engineers for the Next Generation

국가전략기술과 글로벌 경력 개발 - 양자기술을 중심으로 240
박갑동

한계를 넘는 엔지니어의 여정 (부제: 연구실, 군용시스템, 그리고 양자컴퓨터까지) 245
김재인

대학교와 대학원의 차이 (부제: 학사/석사/박사 학위란?) 248
박준범

계획대로 되지 않아도 길은 이어진다. 252
박형원

전임 교원에 대해서 박사과정 때 미리 알았으면 좋았을 것들 256
남세광

Physics Division

Aims

The Physics Division invites voluntary contributions from VeKNI members that reflect the broad and deep research landscape across fundamental and applied physics. Members are highly encouraged to share recent scientific advances, summaries of ongoing research projects, emerging experimental and theoretical trends, technological innovations, or perspectives on future challenges in physical sciences.

Scope

- **Condensed Matter Physics:** Quantum materials, topological phases, magnetism, superconductivity, semiconductors, low-dimensional systems, and advanced functional materials.
- **Quantum Science & Technology:** Quantum information, quantum computing, quantum optics, atom-photon interactions, and the development of quantum devices.
- **Optical Physics & Photonics:** Laser physics, ultrafast spectroscopy, non-linear optics, metamaterials, plasmonics, and advanced optical imaging techniques.
- **Nuclear, Particle & High-Energy Physics:** Fundamental interactions, particle astrophysics, accelerator physics, theoretical high-energy physics, and dark matter searches.
- **Statistical, Nonlinear & Biological Physics:** Complex systems, soft matter, biophysical modeling, non-equilibrium thermodynamics, and fluid dynamics from a physical perspective.
- **Advanced Instrumentation & Computational Physics:** Synchrotron radiation, neutron/X-ray scattering, electron microscopy, first-principles calculations (DFT), molecular dynamics, and machine learning applications in physics.

Monolithic III-V Nanowire Quantum Dot Emitters for Integrated Silicon Photonics

Hyowon Jeong[1,2] and Gregor Koblmüller[1,3]

[1]Walter Schottky Institute, TUM School of Natural Sciences, Technical University of Munich, Garching 85748, Germany
[2]Department of Physics, University of California, Berkeley, Berkeley, California 94720, United States
[3] Institute for Physics and Astronomy, Technical University Berlin, Berlin 10623, Germany
hyowon.jeong@berkeley.edu

Abstract. Integrated quantum photonics requires compact and efficient non-classical light sources that can be site-selectively positioned and incorporated into scalable photonic platforms. Here, we summarize the development of a catalyst-free GaAs-based nanowire (NW) platform for axial quantum-dot (QD) single-photon emitters usingselective-area molecular beam epitaxy.Toovercome limitations in NW growth yield, aspect ratio, and uniformity, dilute Sb was introduced as a surfactant, modifying surface diffusion and thereby improving NW growth while also enabling control over crystal microstructure and carrier recombination dynamics. Building on this platform, optically active axial InGaAs heterostructures were realized and structurally and optically characterized. Automated high-throughput spectroscopy of 16,800 individual NWs further enabled a statistical evaluation of geometry-dependent InGaAsemission yield and bandgap energy. Finally, Sb-mediated suppression of rotational twins was used to control facet-dependent growth and promote confinement of the InGaAs(Sb) insertion to the NW top facet, enabling the formation of few-nanometer-scale axial QDs. Well-localized luminescence and pronounced photon antibunching, with $g^{(2)}(0)$ values consistently below 0.4, confirmed single-photon emission from the InGaAs(Sb) QDs. These results establish a controlled NW-QD growth platform for the development of monolithically integrated quantum light sources.

Keywords: III-V nanowires, quantum dots, selective-area molecular beam epitaxy, single-photon emitters

1. Introduction

Integrated quantum photonics provides a promising route toward scalable quantum information processing, communication, and sensing, but its development critically relies on compact non-classical light sources, such as single-photon emitters, that can be site-selectively positioned and efficiently integrated with photonic circuits [1].Semiconductor quantum dots(QDs) embedded in nanowires(NWs) are particularly attractive candidates for such integrated quantum light sources, offering strong photonic confinement and efficient photon extraction through the tailored NW geometry. Prior theoretical work has demonstrated the potential for efficient coupling of emission from a NW-QD source into a silicon-on-insulator (SOI) waveguide [2], motivating experimental efforts toward the realization of such an architecture. As illustrated in Fig. 1, the broader research roadmap targets the monolithic integration of NW quantum emitters with Si photonic waveguides. In this work, we establish the material and growth platform and ultimately realize axially embedded InGaAs-based QD single-photon emitters in GaAs-based NWs.

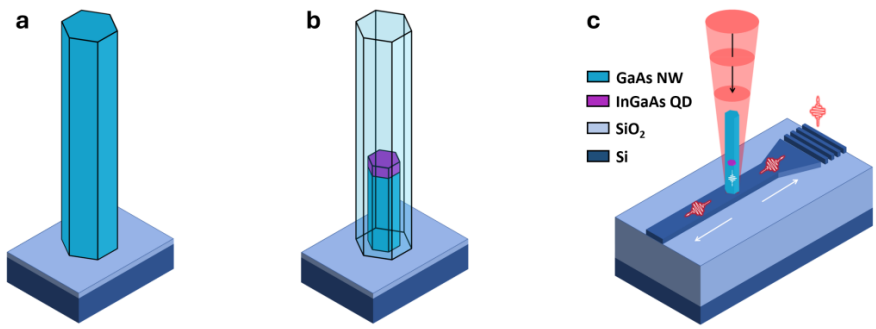


Fig. 1. Research roadmap toward monolithically integrated NW-QDphoton sources: (a) a free-standing GaAs-based NW, (b) an axially embedded InGaAs QD within the NW, and (c) a NW-QD emitter integrated with a silicon-on-insulator waveguide for coupling emitted photons into a photonic circuit. Adapted from Ref. [3].

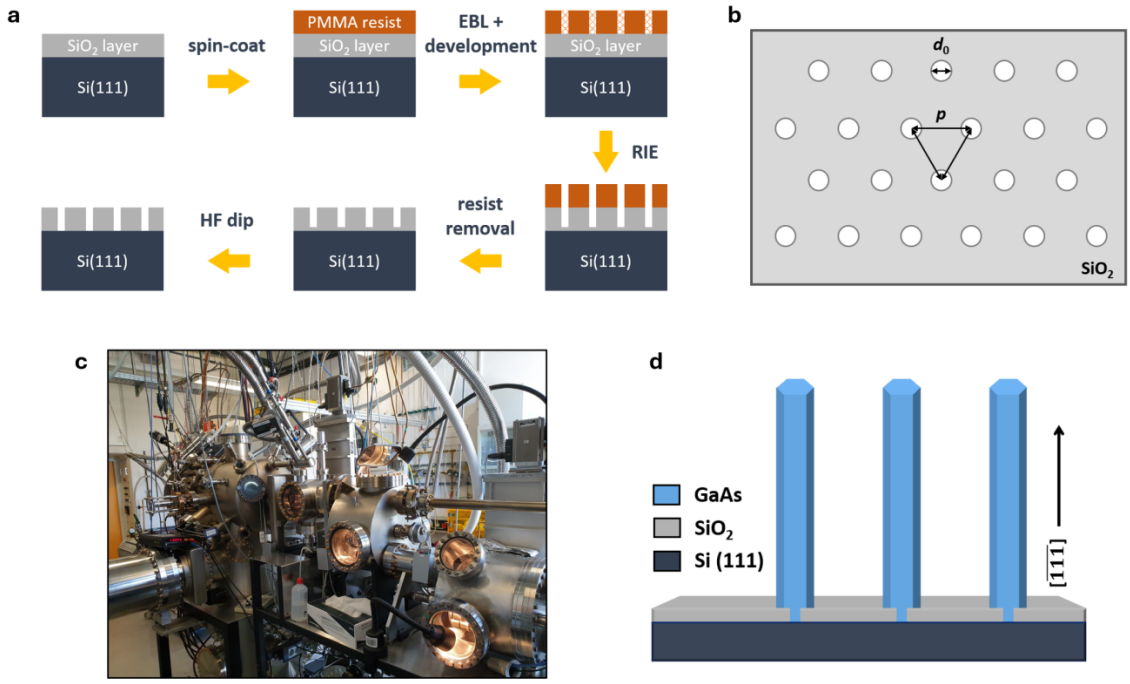


Fig. 2. Site-selective growth of vertical GaAs-based NWs. (a) Fabrication of a patterned SiO_2 mask on Si(111) using electron-beam lithography (EBL)followed by reactive-ion etching (RIE) and hydrofluoric acid (HF) wet etching, (b) a hexagonal array of circular openings formed in the SiO_2 mask, (c) the molecular beam epitaxy system used for NW growth, and (d) vertically aligned NWs selectively grown through the predefined mask openings. Adapted from Ref. [3].

2. Experiments and Discussion

As schematically described in Fig. 2, the GaAs-based NWs were grown by selective-area molecular beam epitaxy, in which atomic fluxes are supplied under ultra-high vacuum to lithographically predefined openings in a dielectric mask. SiO_2-coated Si(111) substrates were patterned by electron-beam lithography followed by dry and wet etching to form hexagonal arrays of circular openings. The opening diameter (d_0) and array pitch (p) provide geometric parameters for controlling the resulting NW dimensions and growth characteristics. Prior to and during growth, thermal treatments, fluxes, and substrate temperature were carefully calibrated to optimize nucleation and adatom diffusion, enabling vertical epitaxial growth along the [111]B direction. In contrast to conventional vapor-liquid-solid growth, the present approach is entirely catalyst-free, thereby avoiding foreign-catalyst contamination and the complex dynamics associated with a catalyst droplet [3,4].

Despite efforts to optimize the growth parameters, catalyst-free GaAs NWs still exhibited limited growth yield, low aspect ratio (length/diameter), and substantial inhomogeneity. Achieving sufficiently long, high-aspect-ratio NWs is particularly important for tuning the resonant cavity length, a key parameter for optimizing the Purcell effect and enabling efficient light out-coupling into photonic waveguides [2]. As shown in Fig. 3(a), the introduction of dilute Sb significantly improved the NW yield and aspect ratio through a surfactant-mediated modification of the surface energetics [4–6]. In particular, the enhanced axial growth indicates that Sb promotes adatom diffusion toward the NW growth front, providing an effective means of controlling NW morphology and growth kinetics.

Beyond this morphological improvement, varying the Sb content within the dilute regime also enabled systematic tuning of the crystal microstructure, particularly the density of rotational twin defects, which locally alter the crystal orientation, as illustrated in Fig. 3(b). These structural changes were further reflected in the charge-carrier recombination dynamics, as shown in the time-resolved photoluminescence (PL) results in Fig. 3(c), demonstrating that dilute Sb can be used to control the morphological, microstructural, and optical properties of the GaAs(Sb) NW platform [6].

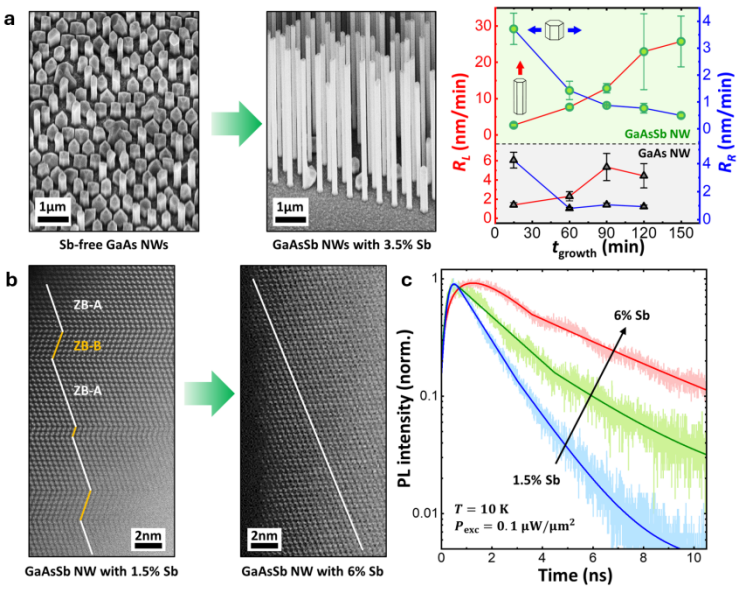


Fig. 3. Establishment and systematic optimization of a catalyst-free GaAs(Sb) NW growth platform using dilute Sb incorporation. Sb-mediated effects on (a) NW morphology and axial/radial growth dynamics, (b) crystal microstructure and twin-defect density, and (c) charge-carrier recombination dynamics measured by time-resolved PL. Reproduced from Ref. [6].

Building on the optimized GaAs(Sb) NW platform, we next introduced a lower-bandgap InGaAs segment along the NW growth axis to form an optically active axial heterostructure. The In content was tuned within a moderate range of ≈10–20%. Since In incorporation is highly sensitive to growth temperature, the temperature was carefully calibrated to maintain sufficient In-adatom diffusion toward the growth front while avoiding significant desorption. The growth time and flux conditions were also adjusted to control the InGaAs segment dimensions. Following the InGaAs insertion, a higher-bandgap AlGaAs passivation shell was grown to suppress non-radiative surface recombination, followed by an outerGaAs shell to protect the Al-containing layer from oxidation [7].

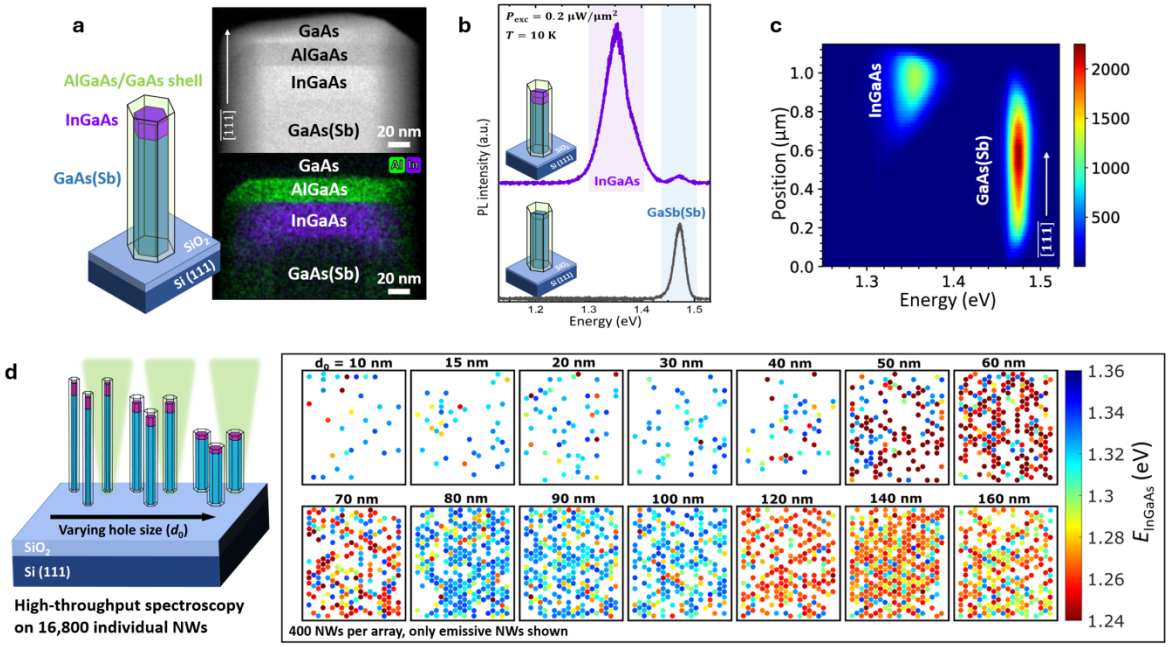


Fig. 4. Development and statistical optical characterization of axial InGaAs NW heterostructures. (a) Structure and compositional characterization of an axially inserted InGaAs active segment. (b) PL and (c) spatially resolved CL measurements confirming emission from the active region. (d) High-throughput spectroscopy of NW arrays for statistical analysis of geometry-dependent optical properties across 16,800 individual NWs. Each colored dot corresponds to a single NW with InGaAs emission, with its bandgap energy indicated by the color scale. Adapted from Refs. [7,8].

The resulting heterostructure is illustrated in Fig. 4(a), demonstrating directional axial incorporation of InGaAswithout undesired radial overgrowth. As shown in Fig. 4(b), NWs containing the InGaAs insertion exhibit an additional emission peak with a lower energy that is absent in reference GaAs(Sb) NWs. Spatially resolved cathodoluminescence (CL) measurements in Fig. 4(c) indicate that this lower-energy emission is well localized to the InGaAsregion, confirming the successful formation of an optically active axial heterostructure [7].

To obtain statistical insight, NW arrays with systematically varied geometrical parameters were investigated using automated high-throughput spectroscopy [8]. In total, 16,800 individual NWs were measured, allowing the luminescence yield and emission-energy trends to be mapped across the geometrical parameter space. Figure 4(d) shows a representative map, where each point corresponds to a NW with detectable InGaAs emission and the color represents the extracted InGaAs bandgap energy. Such mapping provides a direct quantitative view of the array geometries that produce high emission yield, as well as systematic trends in emission energy. A more detailed analysis of these data, including correlations between the emission-energy trends and the underlying NW microstructure, is provided in Ref. [8].

The next step toward a NW-QD emitter was to reduce the axial InGaAs segment to a few-nanometer scale. However, deterministic control of the segment dimensions and composition was limited by the facet geometry of the underlying NW growth front [7]. In particular, rotational twinsstrongly influence the relative growth rates of the top and inclined facets [4], promoting undesired growth on the inclined facets and thereby limiting axial confinement. To overcome this limitation, the Sb-mediated growth control established above for the GaAs(Sb) NW platform [5,6] was extended to the InGaAs insertion [9]. As illustrated in Figs. 5(a,b), dilute Sb suppresses rotational twin formation and stabilizes growth on the top facet while suppressing overgrowth on the surrounding inclined facets. The In compositional maps in Fig. 5(b) further highlight this relationship: ininserts with high twin densities, the In-rich region extends along the inclined facets, whereas in twin-free inserts it remains strongly confined near the top facet. This comparison directly demonstrates the critical role of twin suppression in controlling axial QD formation [9].

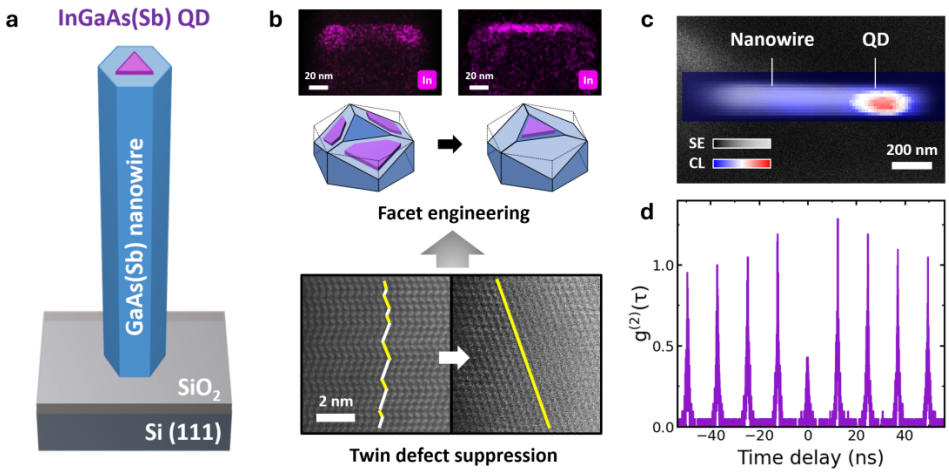


Fig. 5. Realization of axial InGaAs(Sb) QD emitters through Sb-mediated facet engineering. (a) Schematic illustration of the targeted NW-QD structure. (b) Structural micrographs (bottom) and Incompositional maps (top) illustrating the correlation between twin suppression and confined axial growth. (c) CL mapping showing spatially localized emission from the QD region. (d) Second-order photon-correlation measurements exhibiting antibunching with $g^{(2)}(0)<0.4$, confirming single-photon emission. Adapted from Ref. [9].

The optical properties of these NW-QDs were subsequently verified at the single-NW level. The CL map in Fig. 5(c) shows spatially localized emission from the QD region, while second-order photon-correlation measurements in Fig. 5(d) reveal pronounced photon antibunching. In such a measurement, $g^{(2)}(0)$ describes the normalized probability of detecting two photons at zero time delay relative to uncorrelated photon events. While classical Poissonian light gives $g^{(2)}(0)=1$, $g^{(2)}(0)<1$ indicates photon antibunching and non-classical

emission. Values below 0.5 areconventionally regarded as evidence of single-photon emission. The consistently measured $g^{(2)}(0)$ values below 0.4 therefore confirm the single-photon character of the NW-QD emission [9]. Overall, these results demonstrate controlled formation of well-defined axial NW-QD structures and establish the platform as a promising route toward monolithically integrated single-photon sources.

3. Conclusions

In this work, we have summarized the development of a catalyst-free GaAs-based NW platform toward axial QD single-photon emitters. Dilute Sb was first introduced as a self-surfactant, providing effective control over NW growth by enhancing surface diffusion and thereby improving NW yield and aspect ratio, while simultaneously tuning crystal microstructure and carrier recombination dynamics. Based on this optimized platform, optically active axial InGaAs heterostructures were realized, and their geometry-dependent optical properties were further investigated through large-scale high-throughput spectroscopy. The remaining limitation in reducing the InGaAs segment to the QD scale was addressed by extending Sb-mediated growth control to the active insertion. Suppression of rotational twins enabled improved control over facet-dependent growth and axial confinement, resulting in well-defined InGaAs(Sb) QDs with well-localized emission and pronounced photon antibunching. Overall, these advances demonstrate a systematic route from NW growth optimization to single-photon-emitting NW-QDs and provide a material and growth foundation for their future monolithic integration with Si photonic circuits.

References

1. Moody, G., et al.: 2022 roadmap on integrated quantum photonics. Journal of Physics: Photonics 4(1), 012501 (2022).
2. Mukhundhan, N., Ajay, A., Bissinger, J., Finley, J. J., Koblmüller, G.: Purcell enhanced coupling of nanowire quantum emitters to silicon photonic waveguides. Optics Express 29(26), 43068–43081 (2021).
3. Jeong, H.: Towards monolithic III-V nanowire quantum emitters for integrated silicon photonics. Ph.D. thesis, Technische Universität München (2025).
4. Esmaielpour, H., Jeong, H. W., Döblinger, M., Ajay, A., Ruhstorfer, D., Finley, J. J., Koblmüller, G.: A review on vapor-solid growth of GaAs-based nanowires by molecular beam epitaxy: insights from twins, facets, dopants, and dilute species. Nanotechnology, in press (2026). https://doi.org/10.1088/1361-6528/ae987f
5. Ajay, A., Jeong, H. W., Schreitmüller, T., Döblinger, M., Ruhstorfer, D., Mukhundhan, N., Koolen, P. A. L. M., Finley, J. J., Koblmüller, G.: Enhanced growthand properties of non-catalytic GaAs nanowires via Sb surfactant effects. AppliedPhysics Letters 121(7), 072107 (2022).
6. Jeong, H. W., Ajay, A., Yu, H., Döblinger, M., Mukhundhan, N., Finley, J. J., Koblmüller, G.: Sb-mediated tuning of growth- and exciton dynamics in entirely catalyst-free GaAsSb nanowires. Small 19(16), 2207531 (2023).
7. Jeong, H. W., Ajay, A., Döblinger, M., Sturm, S., Gómez Ruiz, M., Zell, R., Mukhundhan, N., Stelzner, D., Lähnemann, J., Müller-Caspary, K., Finley, J. J., Koblmüller, G.: Axial growth characteristics of optically active InGaAs nanowire heterostructures for integrated nanophotonic devices. ACS Applied Nano Materials 7(3), 3032–3041 (2024).
8. Jeong, H. W., Church, S. A., Döblinger, M., Ajay, A., Haubmann, B., Patel, N., Finley, J. J., Parkinson, P. W., Koblmüller, G.: High-throughput spectroscopy of geometry-tunable arrays of axial InGaAsnanowire heterostructures with twin-induced carrier confinement. Nano Letters 24(45), 14515–14521 (2024).
9. Jeong, H. W., Koulas-Simos, A., Limame, I., Döblinger, M., Kim, S. K., Palekar, C. C., Finley, J. J., Reitzenstein, S., Koblmüller, G.: Monolithic axial InGaAs quantum dot emitters in GaAs-based nanowires via Sb-mediated facet engineering. Nano Letters 26(27), 8834–8841 (2026).

Electrical polarization and probe characteristic of ferroelectric van der Waals material α-In_2Se_3

Sunwoong Yang[1], Minwoo Lee[2], Young-Mi Bahk[2], and Sung Ju Hong[1]

[1]Division of Science Education, Kangwon National University, Kangwondaehak-gil 1, 24341 Chuncheon, Republic of Korea
[2]Department of Physics, Incheon National University, Academy-ro 119, 22012 Incheon, Republic of Korea
sjh@kangwon.ac.kr

Abstract. We characterize ferroelectric polarization and resultant current probe feature in 2H α-In_2Se_3 vertical device. The vertical α-In_2Se_3 device is fabricated with metal electrode/α-In_2Se_3/metal electrode geometry and characterized by atomic force microscopy and Raman spectroscopy. Two-terminal electrical measurement exhibits distinct current-voltage characteristic behavior with opposite polarization. The results can be attributed to different band diagram depending on ferroelectric polarization.

Keywords: van der Waals ferroelectric semiconductor, Indium selenide (α-In_2Se_3), Polarization-dependent current-voltage characteristic.

1. Introduction

Two-dimensional (2D) ferroelectric semiconductors have attracted attention as promising building blocks for next-generation electronic device application, due to spontaneous electric polarization and semiconducting properties.[1] Among the various 2D materials, 2H α-phase indium selenide (α-In_2Se_3) has representatively emerged because ofits room-temperature ferroelectricity with stablesemiconducting feature under ambient conditions.[2] Furthermore, interlocking of in-plane and out-of-plane polarization due to displacement of the central Se atom possesses potential for various device configurations; consequently, both vertical and lateral device configurations can be employed to realize ferroelectric polarization switching. [3]

In vertical device configuration, where the ferroelectric α-In_2Se_3 layer is sandwiched between bottom and top electrodes, the interplay between ferroelectric polarization and the resulting band bending at the metal-semiconductor interfaces plays a role in determining the current-voltage (*I–V*) characteristics. Depending on the polarization state, the effective Schottky barrier height can result in asymmetric rectification behavior. Especially, Wan et al. have reported that the rectification behavior using monolayer and few-layer graphene electrodes occurs due to ferroelectric polarization of α-In_2Se_3flake. [2] However, the heterostructuring of graphene and α-In_2Se_3 layer might restrict scalable applications, so an approach compatible with the conventional semiconductor fabrication process is highly needed.

In this study, we investigate electrical polarization and probe characteristics associated with ferroelectric polarization switching in the α-In_2Se_3vertical device. The polarization-direction-dependent current-voltage characteristics are consistent with a previous report in which graphene electrodes were used. Our device architecture possesses typical metal electrodes compatible with the conventional semiconductor fabrication process, implying greater feasibility.

2. Experiments and Discussion

α-In_2Se_3 flake was obtained by mechanical exfoliation from bulk α-In_2Se_3 onto a polydimethylsiloxane (PDMS) substrate. A bottom metal electrode was fabricated by electron-beam lithography, followed by metallization using thermal and electron-beam evaporators (Cr/Au = 3/87 nm). The exfoliated α-In_2Se_3 flake on the PDMS substrate was transferred onto the bottom electrode. After that, a top electrode was fabricated by the same procedure as for the bottom electrode. Atomic force microscopy (Park Systems, Korea) was employed to measure the thickness of the α-In_2Se_3 flake. Raman spectroscopy was performed to identify the α-In_2Se_3 flake. Both characterizations were performed in an ambient environment. For electrical characterization, two-terminal current-voltage behavior of the α-In_2Se_3 vertical device was obtained using a Keithley 4200-SCS semiconductor parameter analyzer (Tektronix, U.S.A.) in a vacuum probe station (WIT, Korea).

Figure 1 shows the geometry of α-In_2Se_3 vertical device, where the schematic side-views (Figures 1(a) and (b)) and optical microscope images (Figures 1(c) and (d)) are presented. As shown in Figures 1(a) and (c), α-In_2Se_3 flake was transferred onto the bottom metal electrode. After that, the top metal electrode was fabricated (Figures 1(b) and (d)). Figure 2(b) shows an atomic force microscope (AFM) image, where the height line profile reveals a thickness of ~25 nm for the α-In_2Se_3 flake. Figure 2(a) exhibits the Raman spectrum of α-In_2Se_3 flake, where the measurement position (green laser spot) is indicated in the inset of Figure 2(a). The representative vibrational modes such as E^1, A_1^1, A_1^2, and A_1^3 are seen near 90, 105, 180, and 195 cm^{-1}, consistent with a previous report on 2H α-In_2Se_3. [4]

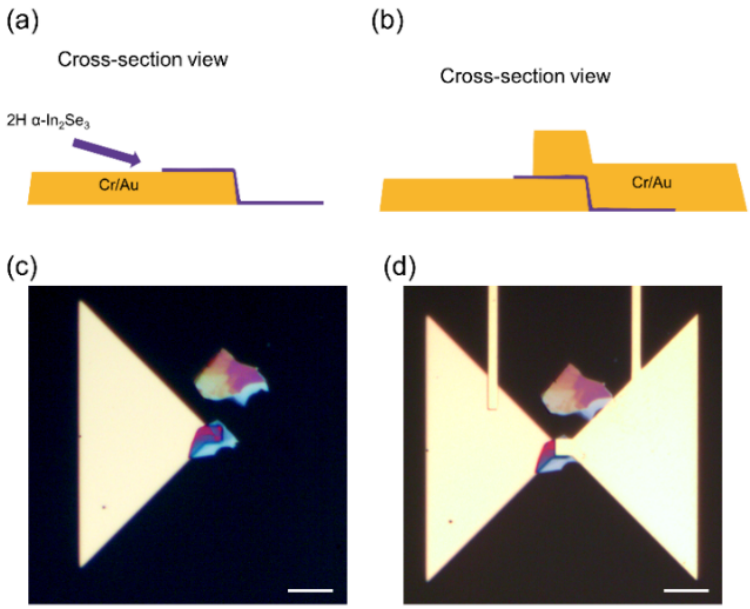


Fig. 1. (a,b) Schematic side-view of 2H α-In_2Se_3 vertical device. (c,d) Optical microscope images. Scale bar is 10 μm.

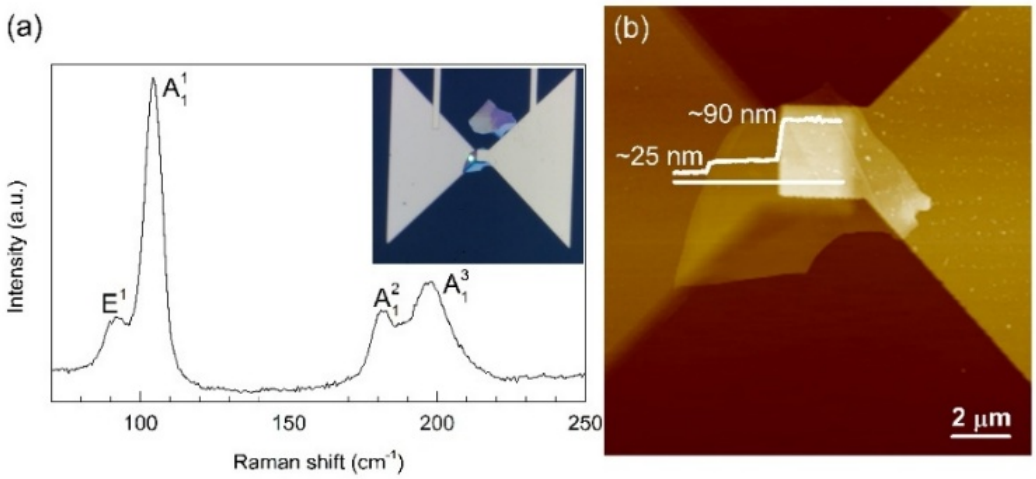


Fig. 2. (a) Raman spectrum of 2H α-In_2Se_3 flake. Inset shows measurement spot with green laser. (b) AFM topography of 2H α-In_2Se_3 flake, where white-line profile exhibits thickness of flake (~25 nm) and metal electrode (~90 nm).

Figure 3 shows the two-terminal *I*–*V*characteristics with different voltage sweeping ranges, where *I* is absolute value of the measured current. As the voltage range increases, more apparent hysteresis behavior occurs: ±1.0 V, ±1.5 V, and ±2.0 V as in Figures 3(a)–(c). For example, minute hysteresis is obtained for the voltage sweeping range of ±1.0 V (Figure 3(a)). In contrast, apparent hysteresis is obtained for the voltage sweeping range of ±2.0 V (Figure 3(c)). The hysteresis can be attributed to electric polarization. In other words, the out-of-plane polarization of α-In_2Se_3 flake formed by the vertical electric field can result in distinct *I*–

*V*characteristic curves, where the polarization-dependent *I*–*V*characteristic is explained by the band diagram with opposite polarization. Namely, different resistance states, a high-resistance state (HRS) and a low-resistance state (LRS), can be formed by a specific polarization direction. [5]

It is noteworthy that the abrupt current change observed in the previous report (Ref. #5) is absent in our study. Even though direct evidence for polarization switching is insufficient, the magnitude of the electric field at a voltage of 2.0 V is higher than the coercive field for polarization switching: namely, thicknesses of ~25 nm (this study) and 120 nm (Ref. #5) were considered at an applied voltage of 2.0 V in the vertical device configuration. Based on the polarization information, we further investigated *I*–*V*characteristic as a probe measurement as shown in Figure 4.

Figure 4 exhibits electrical polarization and probe of the α-In_2Se_3 vertical device. We poled the α-In_2Se_3 vertical device with +2.0 V and –2.0 V, denoted as +poled and –poled, respectively. Once the α-In_2Se_3 vertical device was polarized, an electrical probe measurement was performed within a voltage sweeping range of ±0.6 V, where the voltage sweeping range does not change the polarization direction of α-In_2Se_3 flake. In Figure 4(a), two distinct *I*–*V*characteristic curves are shown, where the blue and red curves were measured in the +poled and –poled states. Depending on the polarization, distinct current is observed. For example, a higher current is obtained at –0.6 V (+0.6 V) in the +poled (–poled) case. The polarization-dependent *I*–*V*characteristic curves are repeatedly obtained: the first and second poled results are presented in Figures 4(a) and (b). Even though the probe features are similar to the results obtained with the graphene/α-In_2Se_3/few-layer graphene vertical device [2], using a conventional metal electrode can offer an advantage in terms of feasibility. Therefore, further basic characterization and investigation are necessary for the metal electrode/α-In_2Se_3/metal electrode vertical device configuration.

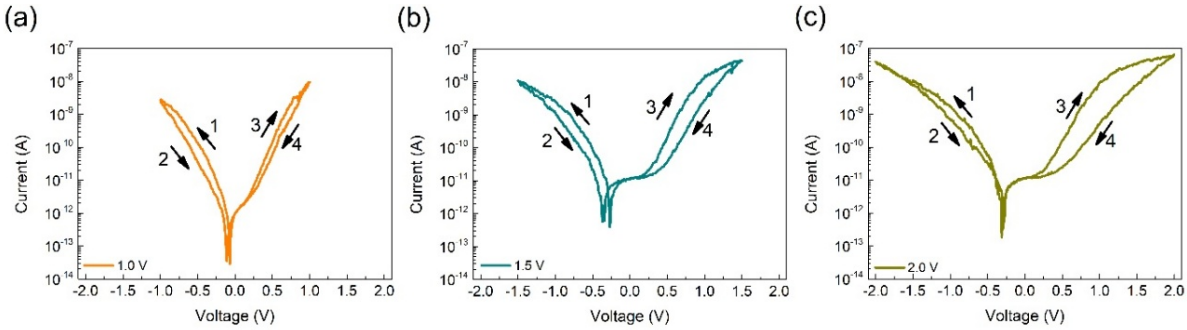


Fig. 3. Current-voltage characteristics of α-In_2Se_3 vertical device with various voltage sweeping ranges: (a) ±1.0 V, (b) ±1.5 V, and (c) ±2.0 V.

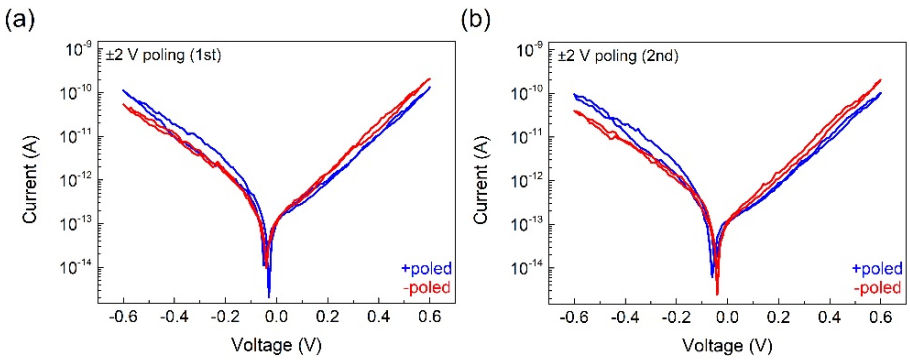


Fig. 4. Electric polarization and probe with poling voltage of ±2.0 V: (a) first and (b) second polarization.

3. Conclusions

In this study, we have investigated electrical polarization and probe ofα-In_2Se_3 with a vertical device configuration. Basic characterizations such as AFM and Raman spectroscopy exhibit the thickness of ~25 nm and typical vibrational modes of α-In_2Se_3.Furthermore, the distinguished *I–V*characteristic was seen, which can be attributed to a polarization-dependent band diagram. The *I–V*characteristic was repeatedly obtained, indicating a robust transport feature through the metal/ferroelectric semiconductor/metal geometry. We expect that the basic electrical characterization of the α-In_2Se_3 vertical device can provide preliminary information for future device applications.

References

1. Si, M., Saha, A.K., Gao, S. et al.: A ferroelectric semiconductor field-effect transistor. Nat. Electron. 2, 580–586 (2019).
2. Wan, S., Li, Y., Li, W. et al.: Room-temperature ferroelectricity and a switchable diode effect in two-dimensional α-In_2Se_3 thin layers, Nanoscale, 10, 14885–14892 (2018).
3. Xiao, J., Zhu, H., Wang, Y. et al.: Intrinsic Two-Dimensional Ferroelectricity with Dipole Locking, Phys. Rev. Lett. 120, 227601 (2018).
4. Kim, J.-H., Kim, S.-H., Yu, H.-Y.: Enhanced Electrical Polarization in van der Waals α-In_2Se_3 Ferroelectric Semiconductor Field-Effect Transistors by Eliminating Surface Screening Charge, Small, 20, 2405459 (2024).
5. Saha, A. K., Si, M., Ye, P. D. et al.: α-In_2Se_3 based ferroelectric-semiconductor metal junction for non-volatile memories, Appl. Phys. Lett. 117, 183504 (2020).

X-ray Photon Correlation Spectroscopy: Probing Nanoscale Dynamics with Coherent X-rays

Wonhyuk Jo[1]

[1] Korea Basic Science Institute (KBSI), Multipurpose Synchrotron Radiation Construction Project, Ochang, Republic of Korea
wins1119@kbsi.re.kr

Abstract. X-ray Photon Correlation Spectroscopy (XPCS) uses the coherent flux of modern synchrotron sources to probe the dynamics of disordered condensed-matter systems on nanometer-to-micrometer length scales and timescales from microseconds to hours, reaching optically opaque or strongly absorbing samples that are inaccessible to dynamic light scattering and systems with unfavorable contrast for neutron spin-echo spectroscopy. Since its first demonstration in the 1990s, XPCS has been applied to colloidal suspensions and gels, metallic and molecular glasses, polymer melts and thin films, and, increasingly, protein and biomolecular dynamics in solution and functional materials under operando conditions. The advent of fourth-generation, diffraction-limited storage rings has increased the coherent flux available for such measurements by one to two orders of magnitude, enabling faster, higher-resolution, and more radiation-sensitive XPCS studies than were previously possible. To provide Korean and international researchers with access to this capability, the ID04 beamline is being constructed as Korea's first synchrotron beamline dedicated to XPCS, as part of the Korea Light Source (KLS). This report summarizes the scientific scope of XPCS and the beamline being built to support it.

Keywords. X-ray Photon Correlation Spectroscopy, Coherent X-ray Scattering, Nanoscale Dynamics, Fourth-Generation Storage Ring, Synchrotron Beamline.

1. Introduction

Understanding how matter rearranges itself at the nanoscale, whether in a colloidal gel aging under gravity, a metallic glass relaxing toward equilibrium, or a protein diffusing in a crowded cellular environment, is central to soft matter physics, materials science, and structural biology alike. Such dynamics typically occur on length scales of nanometers to micrometers and timescales from microseconds to hours, a regime that is difficult to access with conventional structural probes, which capture only static, ensemble-averaged structure.

Dynamic light scattering and neutron spin-echo spectroscopy have long provided access to this regime, but are limited, respectively, to optically transparent samples and to systems with sufficient neutron scattering contrast. X-ray Photon Correlation Spectroscopy (XPCS) overcomes both limitations by using coherent X-rays, which penetrate optically opaque and radiation-dense materials while retaining sensitivity to nanoscale structural fluctuations through the resulting speckle pattern. In this report, we summarize the scientific principles and applications of XPCS, and the beamline currently being constructed in Korea to provide domestic access to this technique.

2. X-ray Photon Correlation Spectroscopy

When a sample with a spatially disordered structure is illuminated with coherent X-rays, the scattered intensity forms a speckle pattern whose fine structure reflects the instantaneous configuration of scatterers within the illuminated coherence volume. As the sample evolves, the speckle pattern fluctuates, and the underlying dynamics can be recovered from the normalized intensity autocorrelation function,

$$g_2(q,\tau) = \langle I(q,t) \cdot I(q,t+\tau) \rangle \,/\, \langle I(q,t) \rangle^2 \qquad (1)$$

whose decay with delay time τ is related to the underlying relaxation process, for example the free-particle diffusion coefficient of a Brownian system via the Stokes-Einstein relation, or a collective diffusion coefficient when interparticle interactions are significant [1].

Equation (1) assumes that the sample dynamics are stationary over the measurement time, so that g_2 depends only on the delay τ and not on the absolute time t. For systems whose dynamics evolve during the measurement, such as an aging gel, a sample undergoing a phase transition, or a material responding to an external stimulus, this assumption breaks down, and the two-time correlation function (TTCF), $C(q,t_1,t_2) = \langle I(q,t_1)\cdot I(q,t_2)\rangle / [\langle I(q,t_1)\rangle\cdot\langle I(q,t_2)\rangle]$, is used instead. Plotted as a two-dimensional map over t_1 and t_2, the TTCF directly visualizes how the sample's relaxation time changes over the course of the experiment, with the ordinary, stationary $g_2(q,\tau)$ recovered by averaging the TTCF along lines of constant $\tau = t_2 - t_1$. This makes the TTCF a particularly valuable tool for non-equilibrium and operando XPCS studies, where the dynamics themselves are often the quantity of scientific interest.

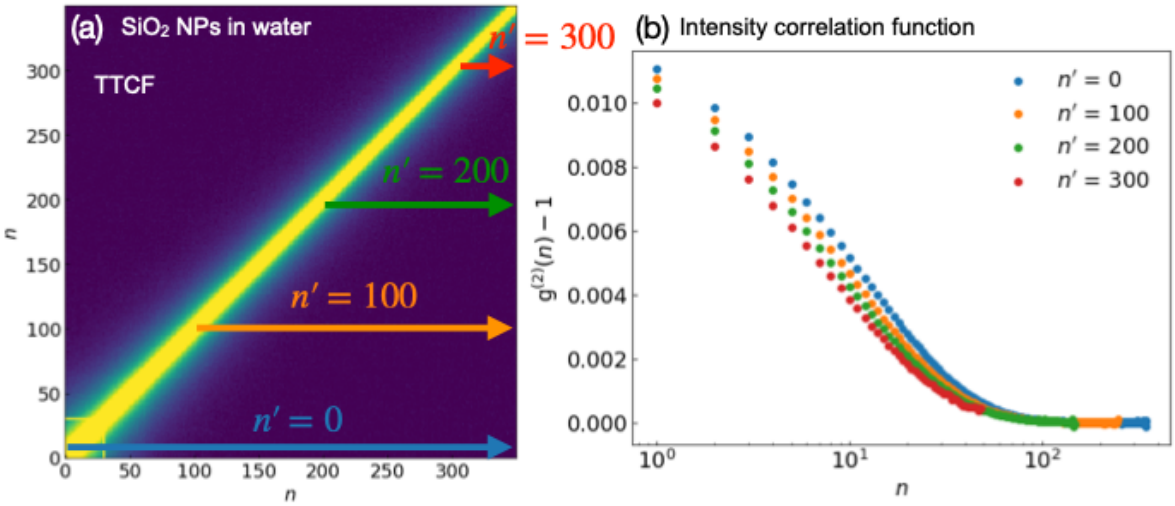


Fig. 1. Example TTCF measurement of SiO_2 nanoparticles in water from the Materials Imaging and Dynamics instrument at the European XFEL. (a) Two-time correlation function, with arrows indicating the extraction of conventional $g_2(q,\tau)$ curves at different numbers of X-ray pulses n′ included in the analysis. (b) Corresponding intensity correlation functions for the indicated values of n′, showing faster decorrelation at larger n′ due to X-ray induced heating.

Since its first experimental demonstration, XPCS has been applied across a broad range of research areas. In soft matter physics, it has been used to follow the aging and yielding of colloidal gels and concentrated suspensions, revealing intermittent, non-diffusive dynamics that are inaccessible to ensemble-averaged scattering. In condensed-matter physics, XPCS has probed the atomic-scale relaxation dynamics of metallic and molecular glasses, including dynamical heterogeneity and aging far from equilibrium. In polymer science, it has characterized the dynamics of polymer melts and thin films under confinement. More recently, XPCS has been extended to protein and biomolecular dynamics in solution [2], and to functional materials such as battery electrolytes and catalysts under realistic operando conditions, tracking nanoscale structural evolution during operation rather than only before and after.

The statistical quality of an XPCS measurement is governed by the speckle contrast β, which depends on the coherence properties of the incident beam, the sample thickness, the scattering geometry, and the detector characteristics. β degrades rapidly once the illuminated beam size or bandwidth exceeds the transverse or longitudinal coherence length, respectively. Third-generation storage rings, based on double- or triple-bend achromat lattices, deliver only a small fraction of their total flux within the coherence volume. Fourth-generation storage rings instead employ multi-bend achromat lattices that reduce the equilibrium emittance by one to two orders of magnitude, approaching the diffraction limit $\varepsilon_r = \lambda/4\pi$ for hard X-rays and correspondingly increasing the coherent flux available for XPCS [3]. This has motivated the deployment or upgrade of dedicated XPCS instruments at essentially every major fourth-generation facility worldwide, including ESRF-EBS, APS-U, and PETRA IV [4].

3. A Dedicated XPCS Beamline in Korea

To provide this capability domestically, the ID04 beamline, dedicated to coherent small-angle X-ray scattering, is being constructed as part of the Korea Light Source (KLS), Korea's first fourth-generation light source, complementing the country's existing third-generation and XFEL facilities at the Pohang Accelerator Laboratory, which do not currently offer a beamline dedicated to XPCS. The beamline is driven by an in-vacuum undulator delivering coherent flux optimized around 9–12 keV, and preserves this coherence through a horizontally diffracting monochromator with a channel-cut crystal for beam-pointing stability, and two diamond Compound Refractive Lens (CRL) transfocators that provide continuously tunable beam size, allowing the speckle contrast and dose delivered to a given sample to be optimized across a wide range of systems. An overview of the beamline layout is shown in Fig. 2.

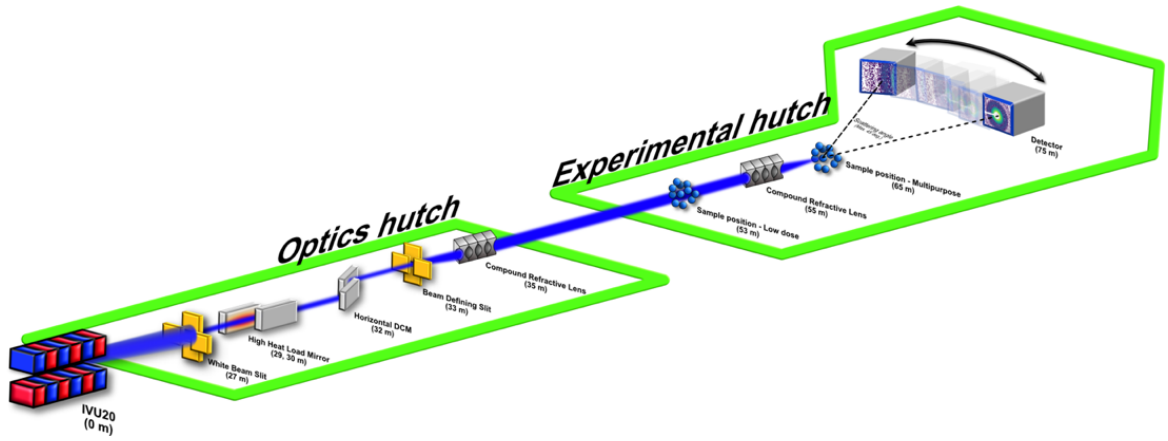


Fig. 2. Overview of the ID04 beamline layout, from the in-vacuum undulator (IVU20) and optics hutch (high-heat-load mirrors, horizontally diffracting monochromator, CRL1) to the experimental hutch (CRL2, Low Dose and Multipurpose sample positions, detector).

The beamline also supports conventional and grazing-incidence Small- and Wide-Angle X-ray Scattering (SAXS/WAXS/GISAXS) [5] as complementary structural probes, and provides two sample positions, a Low Dose position for radiation-sensitive samples and a Multipurpose position for higher-resolution measurements, with an adjustable sample-to-detector distance.

The initial scientific program is expected to draw directly on the application areas described above, including the aging and yielding dynamics of colloidal suspensions and gels, relaxation in metallic and molecular glasses, polymer melt and thin-film dynamics, and, using the Low Dose position, radiation-sensitive protein and biomolecular systems in solution. A dedicated program is also planned for functional and energy materials, such as battery electrolytes and catalysts, studied under in-situ and operando conditions using sample environments including glass capillaries with continuous flow, temperature-controlled cells for time-resolved SAXS, microfluidic devices, and acoustic levitation for container-free measurements of droplets and metastable phases.

The accessible dynamic range is set by the combination of coherent flux and detector time resolution. Slower processes, from milliseconds to hours, such as gelation, aging, and operando structural evolution, are addressed with conventional photon-counting area detectors operated in sequential frame mode. Faster dynamics, from microseconds to milliseconds, relevant to colloidal diffusion and fast relaxation in glasses and liquids, require detectors capable of burst-mode acquisition at rates up to the megahertz range. The beamline is evaluating detector technologies of this class to extend its reach into this regime as they mature.

4. Conclusions

XPCS has become an essential tool for probing nanoscale dynamics across soft matter, condensed-matter physics, polymer science, structural biology, and energy materials research, and is entering a new era of

sensitivity and speed at fourth-generation storage rings. The two-time correlation function further extends this capability to systems whose dynamics are themselves evolving, opening the technique to aging, non-equilibrium, and operando studies that were previously difficult to interpret with conventional, time-averaged analysis.

Once operational, the ID04 beamline will provide the first domestic access to dedicated XPCS measurements, enabling Korean researchers to pursue this class of research without relying on limited international beamtime, and complementing existing Korean synchrotron and XFEL capabilities that are focused on static structure determination. By combining a coherence-preserving optical design with flexible, continuously tunable beam size, the beamline is intended to serve a broad initial user community spanning soft matter, glasses and disordered materials, biomolecular solutions, and functional and energy materials under operando conditions, rather than a single, narrowly defined class of experiment.

Beyond its immediate scientific program, the project is expected to help establish the domestic instrumentation expertise and user community needed to fully exploit coherent scattering techniques as Korea Light Source and future fourth-generation sources come online, and the authors welcome scientific input from the broader community, including VeKNI members, as the design is finalized.

References

1. Grübel, G., Zontone, F.: Correlation spectroscopy with coherent X-rays. J. Alloys Compd. 362, 3–11 (2004).
2. Möller, J., Sprung, M., Madsen, A., Gutt, C.: X-ray photon correlation spectroscopy of protein dynamics at nearly diffraction-limited storage rings. IUCrJ 6, 794–803 (2019).
3. Eriksson, M., van der Veen, J.F., Quitmann, C.: Diffraction-limited storage rings – a window to the science of tomorrow. J. Synchrotron Radiat. 21, 837–842 (2014).
4. Shpyrko, O.G.: X-ray photon correlation spectroscopy. J. Synchrotron Radiat. 21, 1057–1064 (2014).
5. Guinier, A., Fournet, G.: Small-Angle Scattering of X-Rays. Wiley, New York (1955).

Development of periodic Hirshfeld atom refinement

Kanghyun Chu[1]

[1]FS-TUX, DESY, Notkestr.85,22607 Hamburg, Germany
kanghyun.chu@desy.de

Abstract. X-ray diffraction technique has been widely used to investigate microscopic structure of crystal systems. Intensity of each Bragg's peak encodes the information of electron density in the unit cell. X-ray refinement finds the model that fits the measured structure factors the best; atomic coordinates and anisotropic displacement parameters usually comprise the model parameters. Attempts to acquire the quantum mechanical validity of the model gave rise to quantum crystallography. Periodic Hirshfeld atom refinement (pHAR) is developed to rigorously address crystal fields in the molecular electron density calculation. It predicts precise and accurate molecular structures and can refine network compounds or ionic crystals, both of which were inaccessible in the existing HAR framework.

Keywords: X-ray structural determination, Quantum crystallography, periodic Hirshfeld atom refinement

1. Introduction

Crystals are solid materials that are characterized by the spatial periodicity on a microscopic scale. They are found in nature, such as in the ground or asteroids, and they reveal the history and environment since it is formed. Crystals can also be fabricated artificially to study new chemical entities, to invent functional materials, or to explore the exotic quantum world. Single crystals bridge the domains of like physics, chemistry, pharmacy, and materials science.

The periodic nature of crystals makes them work as diffraction gratings. It is an x-ray whose wavelength is comparable to the typical length scale of the internal structures of crystals. Light electrons scatter more incident photons than heavy nuclei do, so the diffracted x-ray beams contain the information about electron density, which also follows the crystal symmetry. The Bragg peak intensity conveys the squared amplitude of the electron density's Fourier component. A model is refined by minimizing the discrepancy between the observed peak intensities and calculated peak intensities. This iterative computational process is called x-ray refinement.

Traditionally, the independent atom model (IAM) has been widely employed to calculate structure factors of a given model. This approach assumes neutral, isotropic, and non-interacting atoms; the electron density of the model crystal is obtained by the superposition of these atoms. The representation of each atom's charge density in reciprocal space—atomic form factors—is tabulated on the International Tables for Crystallography [1]. However, the IAM has an inherent limitation: it does not account for the redistribution of electrons due to chemical bonding and crystal fields are not counted in the model. The atoms in a crystal may be polarized and charged. Assuming them to be spherical and charge neutral causes bias in the model. It is widely known that the bond length involved with hydrogen atoms is underestimated by about 0.1 angstroms. The multipole model [2] attempted to account for electron redistribution by introducing additional free variables to describe the expansion or contraction of the valence orbitals and the multipole moments, at the cost of the risk of having overfitted or physically nonsensical models.

Efforts to create experimental structural models in accordance with quantum mechanics have contributed to the development of quantum crystallography. [3] Hirshfeld atom refinement (**HAR**) has been highly successful in refining structural models by employing the Hirshfeld partitioning scheme and quantum chemical molecular electron density calculations. [4] However, using molecular electron density introduces fundamental limitations. Since the density is calculated from an isolated molecule or molecules, the electron density does not reflect the presence of crystal field. Furthermore, HAR cannot process the network compounds, such as quartz. Taking a fragment out of the crystal always forms broken bonds, so the calculated electron density from the sampled piece is no longer a suitable approximation of the actual electron density in the unit cell. Studies have been conducted to mitigate or overcome the limitations of HAR: classical electrostatic interactions from the neighboring molecules within a certain radius are included in the molecular Hamiltonian to approximate crystal fields [5]; electron density is calculated using the projector-augmented wave method and periodic boundary conditions [6]. Nevertheless, extending the existing HAR to comply with

periodic boundary conditions was desirable.

In this study, we developed periodic Hirshfeld atom refinement (**pHAR**) [7]. This technique is applicable to any periodic system while remaining compatible with conventional HAR by using atom centered Gaussian orbitals with aBloch wave formalism. Compared to the conventional HAR, the refined bond lengths of hydrogen-related bonds showed better agreement with the reference values from neutron diffraction measurements, and the statistical significance of the bond lengths is improved. It also enables one to explore chemical features induced by crystal fields without changing the level of theory.

2. Development of pHAR

Tonto is a software designed for quantum crystallography that can perform HAR. [8] It takes an initial guess of model structure and the measured structure factors or intensities as inputs, and then returns a refined model. It involves repeating cycles of calculating electron density by solving the molecular Hamiltonian, obtaining atomic form factors through partitioning, and updating the parameters using the least squares method to better match the measured structure factors. PHAR is developed by substituting the molecular electron density calculation with the unit cell electron density calculation. When searching for a candidate DFT package, we avoided those using a plane-wave basis or a pseudo-potential; it is required to treat all the electrons, both core and valence, equally, and to ensure the crystal symmetry in the calculated electron density. Crystal23 [9], a program for computational solid state physics and chemistry, aligned with our needs. It uses a Bloch wave basis set, which is made out of atom-centered basis functions, hence satisfying periodic boundary conditions. The flow chart in Figure 1 shows the major differences in the core part of HAR and pHAR.

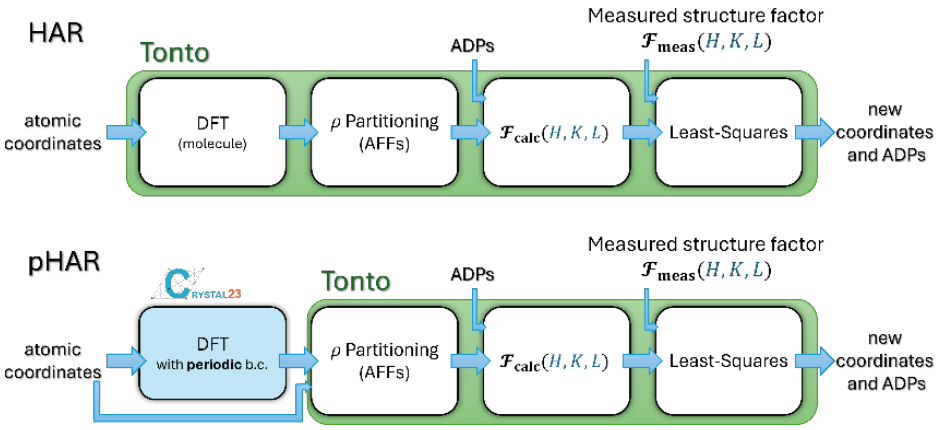


Fig.1. Flow chart of one iteration cycle in the overall process of HAR and pHAR. (ρ stands for electron density. AFFs is atomic form factors. ADPs isatomic displacement parameters. F_{calc} and F_{meas} stand for the calculated and measured structure factors.) It repeats until the starting atomic coordinates and ADPs are unchanged after the process, so that the parameters are considered to be converged.

Tonto uses atom-centered grids, rather than Cartesian grids, to calculate the atomic form factors of each atom. These form factors are obtained via the numerical Fourier transform of the partitioned electron density. This density is given as the grid pointwise product of the unit cell electron density and the Hirshfeld partitioning weight function of the atom of interest. Since Crystal23 provides a function to calculate the electron density only on Cartesian grids, but not at arbitrary points in direct space, a new subroutine was implemented in Tonto. This subroutine calculates electron density on the given grids based on the density matrix from Crystal23 and produces consistent results with those from Crystal23.

Another subroutine was added to enhance the robustness of the detecting symmetry-protected parameters and the stability of the least squares. The calculated structure factors are nonlinear functions of the structural parameters. Therefore, for each iteration, the parameter updates are obtained by calculating the Jacobian matrix J and solving the least squares on the high-dimensional tangent plane in the parameter space. The presence of symmetry-protected parameters makes the matrix $J^T J$ —and consequently the normal equation—to become singular. Tonto finds such constraints by filtering out near-zero eigenvalues of $J^T J$. If the atomic form factors contain numerical errors, meaning they are not invariant under the certain crystal symmetry transformations, then the error propagates through the numerical differentiates and the Jacobian

matrix to reach the eigenvalues that should be zero. This can be prevented by symmetrizing the atomic form factors by using the symmetry operators belonging to the stabilizer subset. This process is equivalent to calculating the average of all possible electron densities, each of which is created by applying each stabilizer to the original imperfect density. (Fig. 2) It was serendipitous that this symmetrization subroutine could enforce the conventional HAR to complete the task without crashing, even when the point group symmetry of certain atoms in the sampled fragment did not align with the local crystal symmetry and the corresponding atomic form factors strongly violated the symmetry invariance condition.

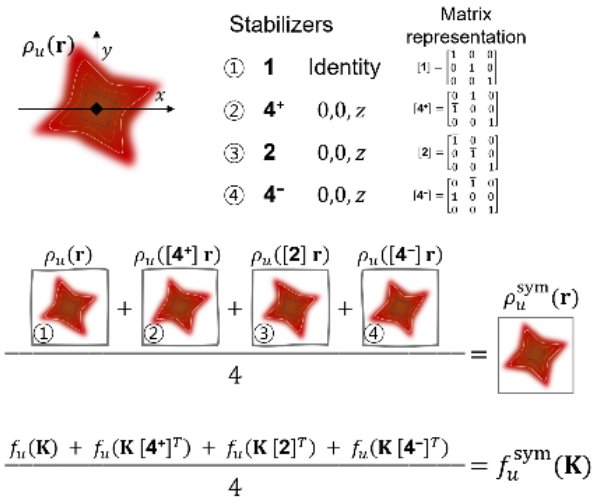


Fig. 2. Schematic representation of the symmetrization of a partitioned atomic electron density sitting on a 4-fold axis. (Adapted from Figure S4 in [7])

The overall iteration control and the convergence test are conducted by lamaGOET [10]. This software provides a graphical user interface to initiate the pHAR autopilot by automatically generating input files for Tonto and Crystal23 for every cycle until a converged structure is achieved.

3. Refinement of ammonia (NH_3) single crystal

We tested pHAR using the X-ray diffraction data of an ammonia single crystal at 160 K. We compared the results refined with pHAR against those from other methods, including IAM, conventional HAR, and HAR with simulated crystal field by using cluster charges (HAR/CC). We kept the same level of theory over HAR, HAR/CC, and pHAR; we used B3LYP hybrid density functional and pob-TZVP-rev2 basis set. These are widely used in the quantum chemistry community and among the Crystal23 users. The refined bond lengths are compared to the reference value by neutron diffraction. (Table 1)

Table 1. R-factor, residual electron density $\Delta\rho$ (in e/Å^3 units) and refined N–H bond length. (Excerpted from Table 1 in [7])

Model	R-factor	$\Delta\rho_{max}$	$\Delta\rho_{min}$	$\Delta\rho_{rms}$	Bond length (Å)
IAM	0.0130	0.0474	-0.0585	0.0130	0.825(8)
HAR	0.0085	0.0310	-0.0233	0.0079	0.933(6)
HAR/CC	0.0078	0.0280	-0.0187	0.0072	0.945(5)
pHAR	0.0069	0.0230	-0.0161	0.0065	0.947(4)
Neutron diffraction					0.989(5)

The R-factors gradually decreases from IAM to HAR to HAR/CC to pHAR. This indicates that the degree of agreement between the model and the measurement increases as we progress from IAM to pHAR. This trend

is also confirmed from the residual electron density $\Delta\rho$. The maximum, minimum, and root-mean-square values also decrease from IAM to pHAR.

As the physical fidelity of crystal field treatment improves, it is observed that the N–H bond length approaches to the neutron reference value. Also, the standard uncertainty gets smaller and smaller, yielding larger statistical significance. However, the N–H bond length predicted by pHAR is improved compared to HAR, but still shorter than the neutron reference by 0.042Å. It suggests that there is significant room for optimizing level of theory.

4. Conclusions

pHAR is a natural extension of HAR, and it can address any crystal, including ones that were previously inaccessible. We compared pHAR with existing HAR techniques. We found that introducing periodic boundary conditions without altering the level of theory was sufficient to improve the accuracy of the refined structure. We anticipate to improve the utility of pHAR through further optimization of the basis set and the theoretical model used to construct the Hamiltonian.

References

1. International Tables for Crystallography vol. C: Mathematical, physical and chemical tables. edited by Prince, E. International Union of Crystallography (2006).
2. Hansen, N. K., Coppens, P.: Testing aspherical atom refinements on small-molecule data sets. Acta Crystallogr. Sect. A **34**, 909-921 (1987).
3. Grabowsky, S., Genoni, A., Bürgi, H.-B., Quantum crystallography. Chem. Sci. **8**(6), 4159-4176 (2017).
4. Capelli, S.C., Bürgi, H.-B., Dittrich, B., Grabowsky, S., Jayatilaka, D.: Hirshfeld atom refinement, IUCrJ**1**,361-379 (2014).
5. Fugel, M. *et al.*, Probing the accuracy and precision of Hirshfeld atom refinement with HARt interfaced with Olex2, IUCrJ**5**, 32-44 (2018).
6. Ruth, P. N., Herbst-Irmer, R., Stalke, D., Hirshfeld atom refinement based on projector augmented wave densities with periodic boundary conditions, IUCrJ**9**, 286-297 (2022).
7. Chu, K., *et al.*, Periodic Hirshfeld atom refinement, J. Phys. Chem. Lett. **17** (11), 3170-3179 (2026).
8. Jayatilaka, D., Grimwood, D. J.: Tonto: A Fortran Based Object-Oriented System for Quantum Chemistry and Crystallography, Computational Science-ICCS 2003**2660**, 142-151 (2003).
9. Erba, A. *et al.*: CRYSTAL23: A Program for Computational Solid State Physics and Chemistry, J. Chem. Theory Comput. **19** (20), 6891–6932 (2023).
10. Malaspina, L.A., Genoni, A., Grabowsky, S.: *lamaGOET*: an interface for quantum crystallography, J. Appl. Cryst. **54**, 987-995 (2021).

Recent Advances in Single-Particle Imaging at XFEL Facilities

Chan Kim

European XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany
chan.kim@xfel.eu, kimchan8237@gmail.com

Abstract. The development of X-ray free-electron lasers (XFELs) has opened new opportunities for single-particle imaging (SPI) by providing ultrashort, highly coherent, and extremely intense X-ray pulses for coherent diffraction measurements from individual particles without crystallization. Advances in XFEL sources, sample delivery, detectors, and computational methods have significantly expanded biological SPI for studying viruses, macromolecular assemblies, and other non-crystalline specimens. The unique properties of XFELs have also enabled SPI to move beyond static structural characterization. In particular, the combination of optical pump-XFEL probe techniques has led to time-resolved SPI, providing a powerful approach for visualizing ultrafast structural dynamics in functional nanomaterials and, ultimately, biological systems.

Keywords: Single particle imaging, time-resolved imaging, functional materials, localized surface plasmon, X-ray free-electron laser, European XFEL

1. Introduction

X-ray free-electron lasers (XFELs) [1-3] have opened new possibilities for imaging matter with high spatiotemporal resolution. Their extremely high peak brightness and coherence enable the collection of coherent diffraction signals from individual non-crystalline particles before radiation damage destroys the sample (diffraction-before-destruction) [4]. This capability forms the basis of single-particle imaging (SPI), in which a series of two-dimensional coherent diffraction patterns from randomly oriented particles are collected and subsequently combined using the expand-maximize-compress algorithm for orientation recovery in reciprocal space [5,6], followed by phase retrieval [7,8] to reconstruct a three-dimensional structure in real space. Unlike conventional crystallography or serial femtosecond crystallography [9,10], SPI does not require the formation of crystals, making it particularly attractive for studying heterogeneous or non-crystallizable specimens.

Much of the early development of SPI has focused on biological specimens, including viruses, proteins, and macromolecular assemblies [11-13]. Considerable progress has been made in sample injection, diffraction-pattern classification, orientation determination, three-dimensional diffraction-volume reconstruction, and phase retrieval, allowing increasingly detailed structural information to be extracted from large numbers of individual particles. In particular, developments in high-repetition-rate XFEL operation [3] and detector technology [14] have enabled the collection of substantially large datasets, which is essential for overcoming the inherently weak diffraction signals from single particles. These continuing developments are steadily expanding the applicability of SPI and are providing a foundation for moving beyond static structural characterization toward studies of structural dynamics.

By combining the ultrashort and intense XFEL pulses with external stimuli, such as an optical pump laser, time-resolved SPI (TR-SPI) can capture transient structural changes over a wide range of timescales, from femtoseconds to microseconds. This approach is particularly promising for functional nanomaterials, where optical excitation can induce ultrafast processes such as lattice deformation, acoustic vibrations, phase transitions, and structural rearrangements. TR-SPI provides the possibility of resolving structural responses at the level of individual particles and investigating variations arising from particle size, shape, orientation, or structural heterogeneity. These developments suggest that SPI can evolve from a technique primarily aimed at determining static structures into a broader imaging approach for studying nonequilibrium structural dynamics.

2. Single-Particle Imaging Based on Coherent X-ray Diffraction Imaging

Figure 1 shows a schematic diagram of the SPI experimental setup. Sample particles, assumed to have nearly identical structures, are continuously delivered to the beam-sample interaction point using a particle injector. At the same time, X-ray pulses are delivered to the interaction point, and the diffraction signals generated when the X-rays interact with individual particles are repeatedly measured. When a particle injector is used, the particles arrive at the interaction point in random orientations, requiring the missing orientation information to be recovered from the measured diffraction patterns.

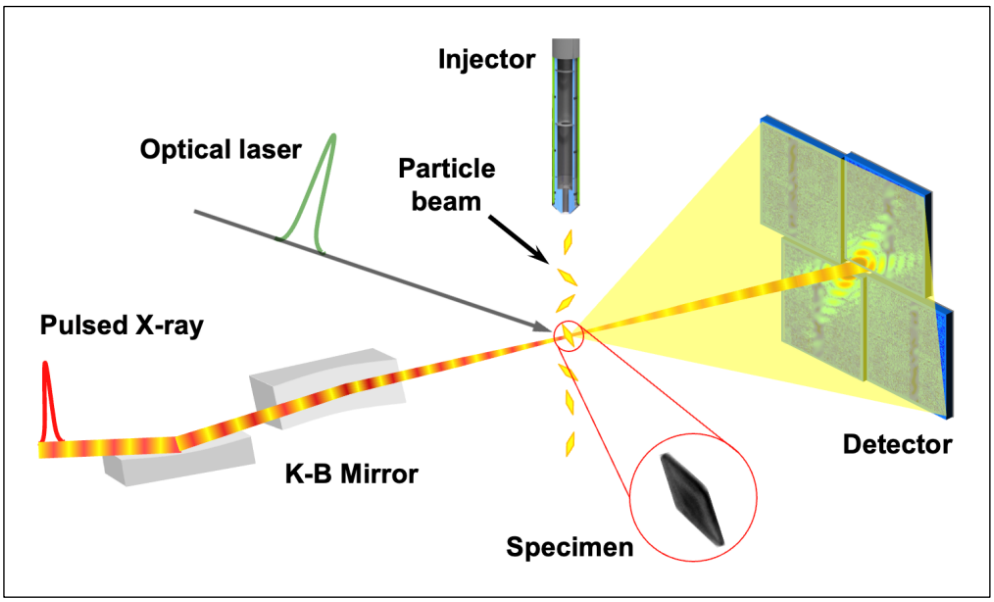


Fig.1. Schematic illustration of a SPI experiment: XFEL pulses illuminate individual particles delivered to the interaction point by a particle injector. Each coherent diffraction patten from a randomly oriented particle is recorded by a two-dimensional detector. A three-dimensional structure is then reconstructed from the collected diffraction patterns through orientation determination and phase retrieval processes.

The data analysis procedure is summarized in Fig. 2. The process starts from collecting a large number of two-dimensional diffraction patterns. And then, the expand-maximize-compress (EMC) algorithm [5,6] is used to estimate the missing orientation information of each diffraction pattern and, based on these orientations, reconstruct a three-dimensional diffraction intensity distribution (Fig. 2(a) and (b)). Finally, a phase retrieval algorithm is applied to reconstruct a three-dimensional real-space image, which contains information about the electron density distribution of the sample (Fig. 2(c) and (d)) [15].

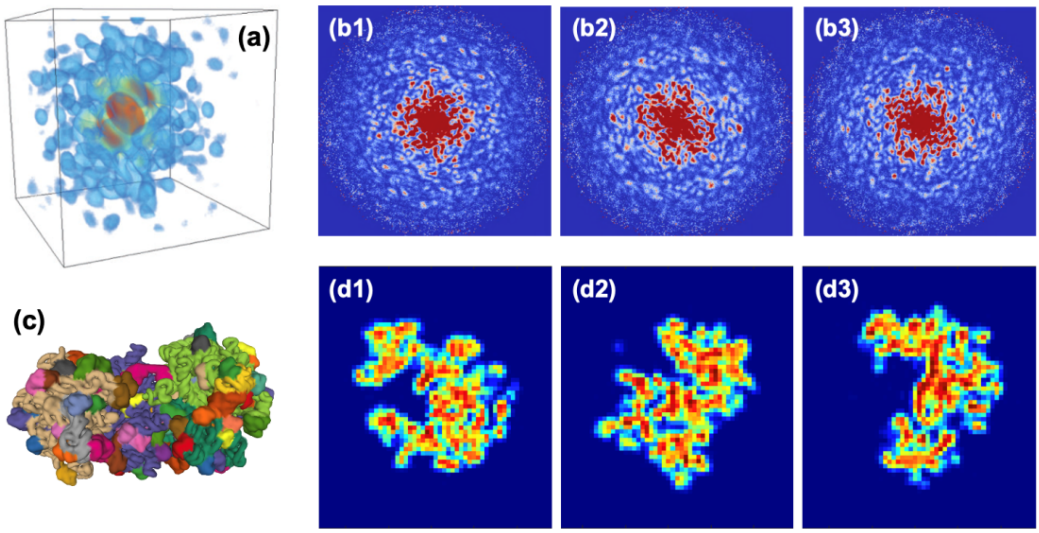


Fig. 2. Three-dimensional image reconstruction process in SPI: A large number of two-dimensional diffraction patterns are collected from nearly identical particles in random orientations. (a) The Expand-Maximize-Compress (EMC) algorithm is used to determine the particle orientations and assemble a 3D diffraction volume. (b1-b3) Central slices of the reconstructed 3D diffraction volume in the yz (b1), xz (b2), and xy (b3) planes. (c) 3D real-space electron density reconstructed by a phase retrieval algorithm. (d1-d3) Central slices of the reconstructed 3D electron density in the yz (d1), xz (d2), and xy (d3) planes. Figure partially adapted from [Ref. 15: J. E. *et al.*, Structural Dynamics **9**, 064101 (2022)].

Early experimental demonstrations of biological SPI were performed at X-ray free-electron laser (XFEL) facilities using isolated viruses and other macromolecular particles [11-13], establishing the feasibility of collecting diffraction signals from individual non-crystalline biological specimens. These pioneering experiments demonstrated that single-particle diffraction patterns could be recorded and combined to obtain three-dimensional structural information without the need for crystallization. Subsequent advances in computational methods, including SPI reconstruction without symmetry constraints [16] and correlation-based analysis [17], improved particle orientation determination and the performance of EMCand phase-retrieval procedures. Together, these experimental and computational developments established the foundation for increasingly reliable three-dimensional structural characterization of individual biological particles at XFEL facilities.

3. Time-Resolved Single-Particle Imaging

Time-resolved single-particle imaging (TR-SPI) is an imaging technique that uses an external pump source to stimulate a sample and an X-ray free-electron laser (XFEL) pulse to monitor the resulting structural changes as a function of time. Ultrashort optical laser pulses are commonly used as the pump source, while electric fields, magnetic fields, temperature changes, or reactive gases can also be employed depending on the experimental requirements. Because structural changes induced by optical laser excitation can occur rapidly, ranging from femtosecond to microsecond timescales, ultrashort XFEL X-ray pulses can be used as probes to record diffraction signals before significant radiation damage occurs. In these experiments, the time delay between the pump and probe pulses is precisely controlled to collect structural information from nominally identical single particles at different stages of structural evolution. This approach enables the temporal evolution of pump-induced structural changes to be reconstructed in two or three dimensions.

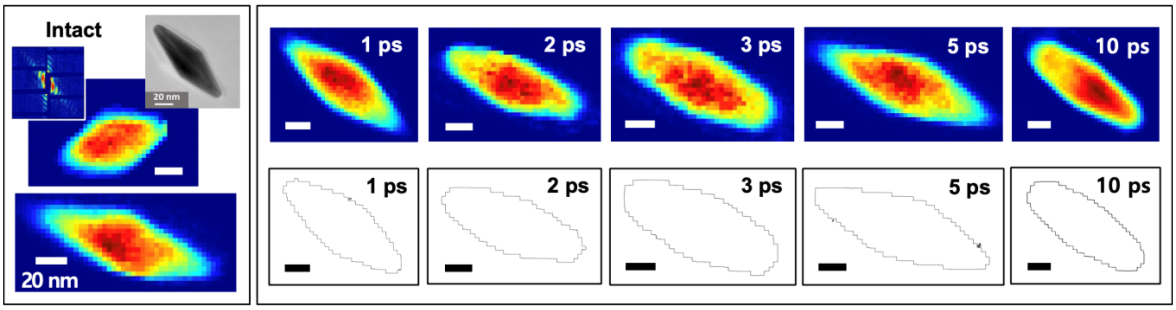


Fig. 3. Representative results of time-resolved SPI: (Left panel) TEM image, representative 2D diffraction pattern, and projected 2D electron-density maps of the Au bipyramid nanoparticles. (Right panel) Reconstructed acoustic modes of the Au bipyramid nanoparticles following femtosecond optical laser excitation. An 800 nm femtosecond optical laser with a fluence of 10 mJ/cm^2 was used to excite the surface electrons. The upper row shows the reconstructed 2D phasing results as a function of pump-probe delay time ranging from 1 ps to 10 ps. The lower row shows the corresponding projected surface area of the reconstructed images, highlighting the ultrafast acoustic vibrations of the nanoparticle.

Figure 3 shows a representative TR-SPI measurement of Au bipyramidal nanoparticles (Au BPNPs) excited by an 800 nm femtosecond laser at a fluence of 10 mJ/cm^2. Coherent X-ray diffraction patterns were collected before and after laser excitation, with time delays from 1 to 10 ps. The left panel shows a TEM image (upper right corner), a representative diffraction pattern (upper left corner), and phase-retrieved real-space images resolving the characteristic nanoparticle facets. The upper row of the right panel presents representative two-dimensional reconstructions at each delay, while the bottom row shows the corresponding particle contours to highlight the localized surface plasmon resonance (LSPR)-induced shape deformation. At 1 ps, the contact angle between neighboring facets along the major axis increases slightly, followed by a more ellipsoidal shape at 3 ps. The particle partially recovers its initial morphology at 5 ps before becoming more ellipsoidal again at 10 ps.

These observations reveal a transient and oscillatory structural response of the Au BPNPs on the picosecond timescale. The initial deformation is likely associated with the LSPR-induced excitation of the nanoparticle, which produces a spatially localized response and drives anisotropic deformation along the particle's major axis. The subsequent partial recovery and renewed deformation suggest that the structural dynamics are not governed by a simple monotonic relaxation process, but instead involve coupled electronic, thermal, and mechanical responses. The ability of TR-SPI to capture these transient morphological changes from individual

nanoparticles provides direct insight into ultrafast nonequilibrium structural dynamics that may be obscured in ensemble-averaged measurements.

4. Conclusion

Coherent X-ray diffraction–based single-particle imaging provides a powerful approach for determining the structure of individual non-crystalline particles without the need for crystallization. By combining randomly oriented diffraction patterns with orientation-recovery and phase-retrieval algorithms, three-dimensional structural information can be reconstructed from a large number of single-particle measurements. Extending this approach to the time domain through TR-SPI enables pump-induced structural dynamics to be investigated with ultrafast temporal resolution. The Au bipyramidal nanoparticles example demonstrates that transient structural deformations and acoustic responses can be directly visualized on the picosecond timescale, providing access to nonequilibrium dynamics that are difficult to resolve using conventional ensemble-averaged techniques. However, three-dimensional time-resolved SPI remains experimentally and computationally challenging. In addition to the inherently low diffraction signal from individual particles, especially for biological specimens, a sufficiently large number of high-quality diffraction patterns must be collected at each pump–probe delay to recover the particle orientations and reconstruct a reliable three-dimensional diffraction volume. Maintaining consistent sample delivery, pump excitation, particle-to-particle reproducibility, and temporal synchronization further complicates data acquisition. Moreover, the structural evolution of individual particles can introduce additional heterogeneity, making orientation determination and 3D reconstruction more difficult than in static SPI. These challenges become particularly significant when extending measurements from two-dimensional projections to fully three-dimensional time-resolved structures. Continued improvements in XFEL source performance, sample delivery, pump–probe synchronization, detectors, and reconstruction algorithms will therefore be essential for realizing routine 3D TR-SPI and enabling direct visualization of ultrafast structural dynamics in increasingly complex nanoscale systems.

References

1. Emma, P. *et al.*: First lasing and operation of an ångstrom-wavelength free-electron laser. Nature Photonics **4**, 641-647 (2010).
2. Kang, H.-S. *et al.*: Hard X-ray free-electron laser with femtosecond-scale timing jitter. Nature Photonics **11**, 708-713 (2017).
3. Decking, W. *et al.*: A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator. Nature Photonics **14**, 391-397 (2020).
4. Neutze, R. *et al.*: Potential for biomolecular imaging with femtosecond X-ray pulses. Nature **406**, 752-757 (2000).
5. Loh, N.-T. D. *et al.*: Reconstruction algorithm for single-particle diffraction imaging experiments. Physical Review E **80**, 026705 (2009).
6. Ayyer, K. *et al.*: *Dragonfly*: an implementation of the expand-maximize-compress algorithm for single-particle imaging. Journal of Applied Crystallography **49**(4), 1320-1335 (2016).
7. Fienup, J. R.: Reconstruction of an object from the modulus of its Fourier transform. Optics Letters **3**(1), 27-29 (1978).
8. Fienup, J. R.: Phase retrieval algorithms: a comparison. Applied Optics **21**(15),2758-2769 (1982).
9. Grünbein, M. L. *et al.*: Megahertz data collection from protein microcrystals at an X-ray free-electron laser. Nature Communications **9**, 3487 (2018).
10. Wiedorn, M. O. *et al.*: Megahertz serialcrystallography. Nature Communications **9**, 4025 (2018).
11. Seibert, M. M. *et al.*: Single mimivirus particles intercepted and imaged with an X-ray laser. Nature **470**, 78-81 (2011).
12. Gallagher-Jones, M. *et al.*: Macromolecular structures probed by combining single-shot free-electron laser diffraction with synchrotron coherent X-ray imaging. Nature Communications **5**, 3798 (2014).
13. Ekeberg, T. *et al.*: Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-Ray Free-Electron Laser. Physical Review Letters **114**, 098102 (2015).
14. Allahgholi, A. *et al.*: The Adaptive Gain Integrating Pixel Detector at the European XFEL. Journal of Synchrotron Radiation **26**(1), 74-82 (2019).
15. E, J. *et al.*: Expected resolution limits of x-ray free-electron laser single-particle imaging for realistic source and detector properties. Structural Dynamics **9**(6), 064101 (2022).
16. Rose, M. *et al.*: Single-particle imaging without symmetry constraints at an X-ray free-electron laser. IUCrJ**5**(6), 727-736 (2018).
17. Kurta, R. P. *et al.*: Correlations in Scattered X-Ray Laser Pulses Reveal Nanoscale Structural Features of Viruses. Physical Review Letters **119**, 158102 (2017).

Environmental Liquid Cell for Three-Dimensional Nanoscale Control of Liquid Water

Chiwon Lee

Verein Koreanischer Naturwissenschaftler und Ingenieure e.V., Konrad-Zuse-Straße, 44801Bochum Süd, Germany
chiwonlee85@gmail.com

Abstract. We demonstrate the formation of continuous films and mesoscopic droplets of liquid water by exploiting the environmental liquid cell technique. Transmission electron microscopy is used for real-time monitoring of the liquid water condensates forming on an electron-transparent membrane. Our quantitative analysis reveals different condensation dynamics between hydrophilic-treated and non-treated surfaces, providing insights into active, three-dimensional control of nanometer-scale liquid water.

Keywords: liquid-phase transmission electron microscope, environmental liquid cell, condensation

1. Introduction

The molecular arrangement of liquid water is one of the most recurring and debated scientific research themes over the past decades [1-3]. The general consensus lies in a tetrahedral configuration where each H_2O molecule participates in four hydrogen bonds, driving the formation of short-range correlations known as the hydrogen bond network [4]. Beyond this classic view, however, a comprehensive theory linking the microscopic structure and unique macroscopic properties of liquid water remains elusive, especially under extreme thermodynamic conditions. Although various total scattering measurements using synchrotrons, X-ray free electron lasers (XFEL), and neutron accelerator have successfully mapped a broad range of temperatures and pressures across the water phase diagram, the resulting real-space structural models have posed questions within the relevant scientific community [5,6].

High-energy free electrons represent an effective alternative probe for atomic structure, offering sub-Ångström wavelengths and large elastic scattering cross-sections driven by electrostatic Coulomb interactions with matter. In transmission electron microscopy (TEM), electron beams enable both real-space imaging and diffraction measurements [7], thus providing atomic scale insight into liquid water. However, the fundamental challenge in structural studies with electron probes stems from isolating and controlling nanoscale liquid water within the high-vacuum column of the microscope. To overcome this limitation, various encapsulation and delivery strategies, including graphene liquid cell [8], direct liquid injection [9], capillary force-driven nanochannel [10], have been developed– each presenting advantages and trade-offs. In this study, we use environmental liquid cell (ELC) technology for the active control of the condensate water on the order of 100 nm in all three dimensions.

2. Experiments and Discussion

Figure 1 briefly describes the operational principle of ELC. Two silicon-based chips, one for the bottom and the other for the top, are assembled and hermetically sealed onto a custom-made TEM holder. Each chip features a U-shaped microchannel with a 25 nm thin rectangular-shaped (25x200 μm) silicon nitride (SiN) freestanding membrane at its middle, used as an observation window under the microscope. Humid air is circulated through the channel via the inlet and outlet port of the bottom cell. By regulating the flow rate and inlet pressure, water condensation initiates, forming a thin liquid layer or droplet at nanoscale, depending on the surface characteristic of the membrane. We test this ELC concept on two different types of membranes: one with a hydrophilic (Figure 2) and a bare SiN (Figure 3) membrane. The hydrophilic membrane is prepared with an oxygen plasma treatment after drop-casting and drying citrate-capped gold nanoparticle solution for enhancing image contrast. Figure 2 shows time-resolved electron micrographs of water condensates that grow and eventually fill the entire observation window within approximately half a minute, resulting in a continues liquid water film on the hydrophilic surface. The estimated liquid thickness is 100±50 nm based on Lambert-Beer's law derived from the ratio between the transmitted and incident beam intensity. In contrast, on a bare SiN membrane, liquid water droplets ranging from 100 - 1500 nm are formed just as dewetting of macroscale

water on raincoats or leaves of plants. We observe time-dependent micrographs of two zeptoliter water droplets in close proximity (~30 nm) on the membrane, by increasing the flow rate of the humid air to induce the drying process (Figure 4). The respective droplets initially approach each other and coalesce into a larger droplet within ~30 sec. Subsequently, a smaller bud emerges on the upper part, liberating from the mother droplet, and eventually evaporating.

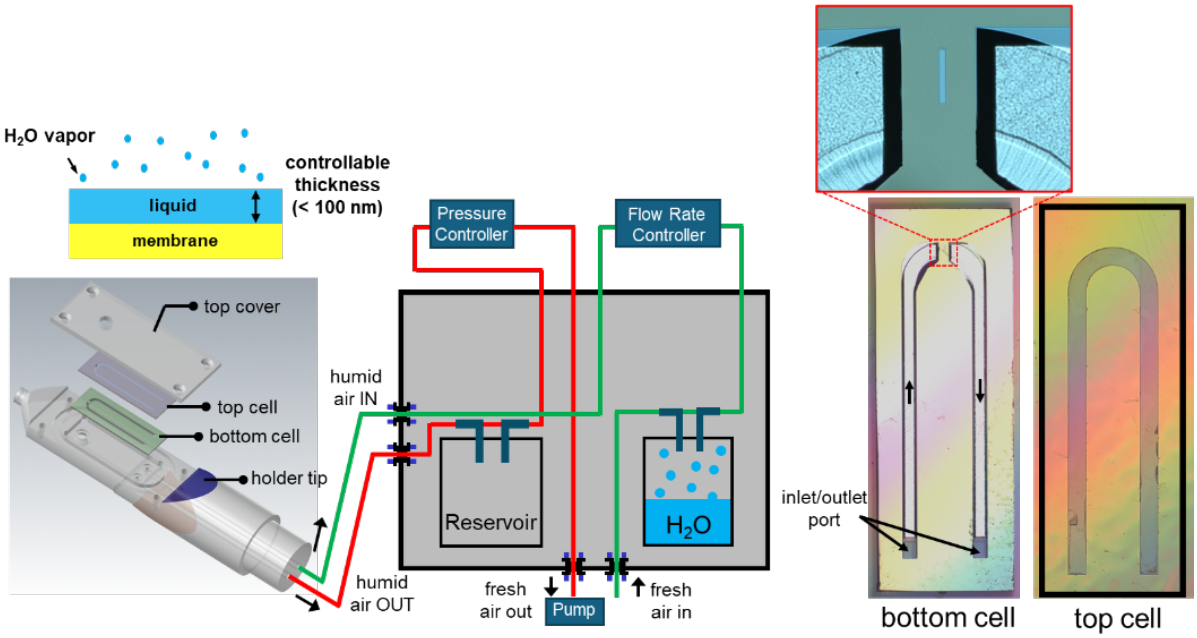


Fig. 1. Schematic illustration of environmental liquid cell concept and cell structure showing a flow channel and SiN observation window.

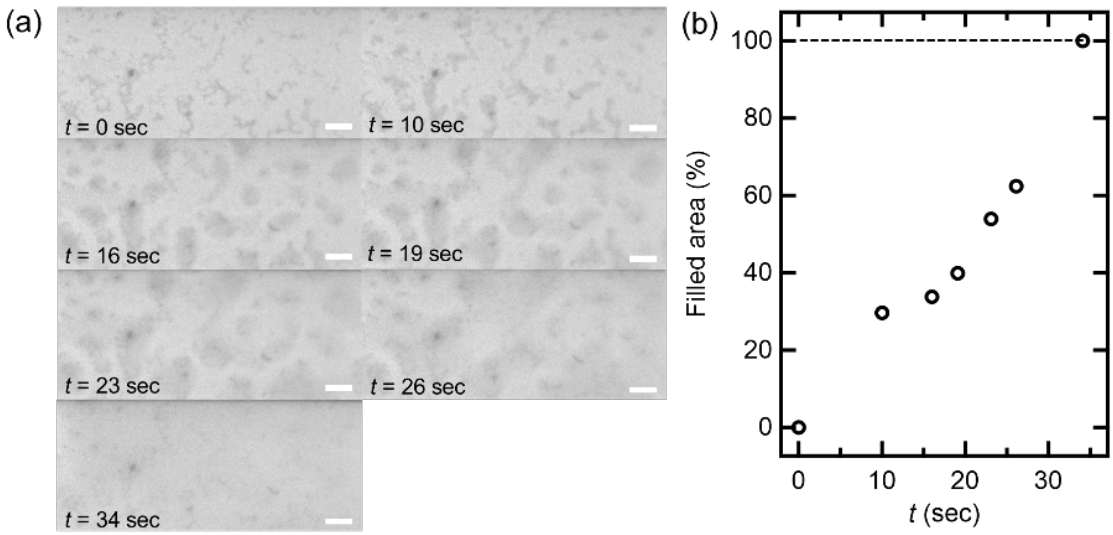


Fig. 2. Formation of 100 nm thick continues liquid water layer on a hydrophilic surface. (a) TEM images captured as a function of time upon humid air flow rate control. (b) Evaluated filled area with liquid water on the observation field of view. Scale bar is 1 μm.

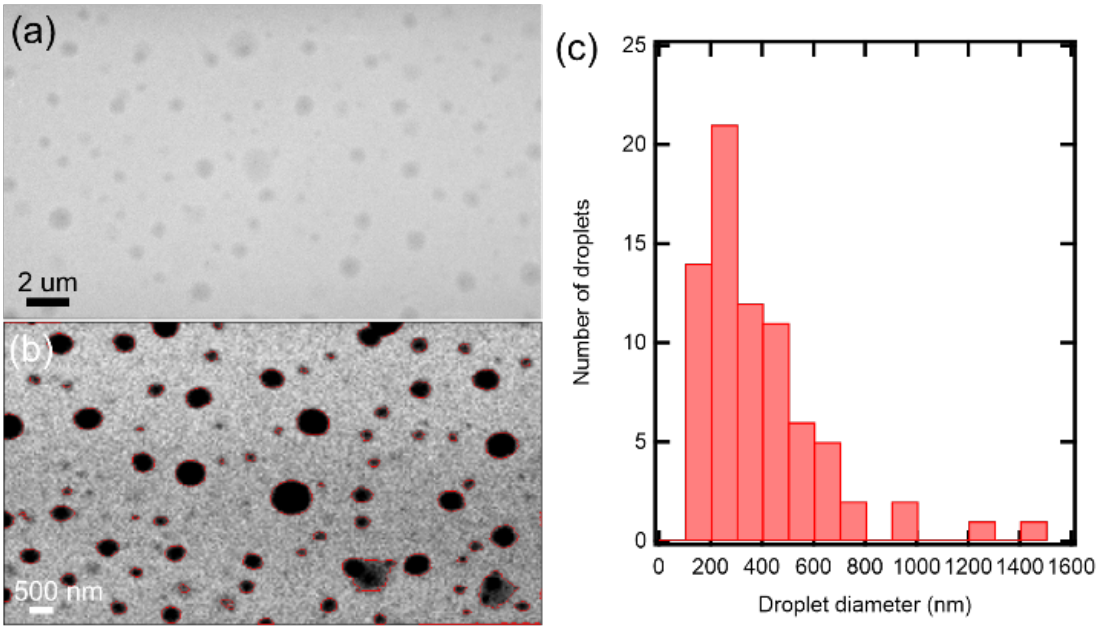


Fig. 3. Formation of zeptoliter liquid nanodroplet on a hydrophobic surface. Typical real-space TEM (a) and STEM (b) images. (c) Droplet size distribution on the observation field of view.

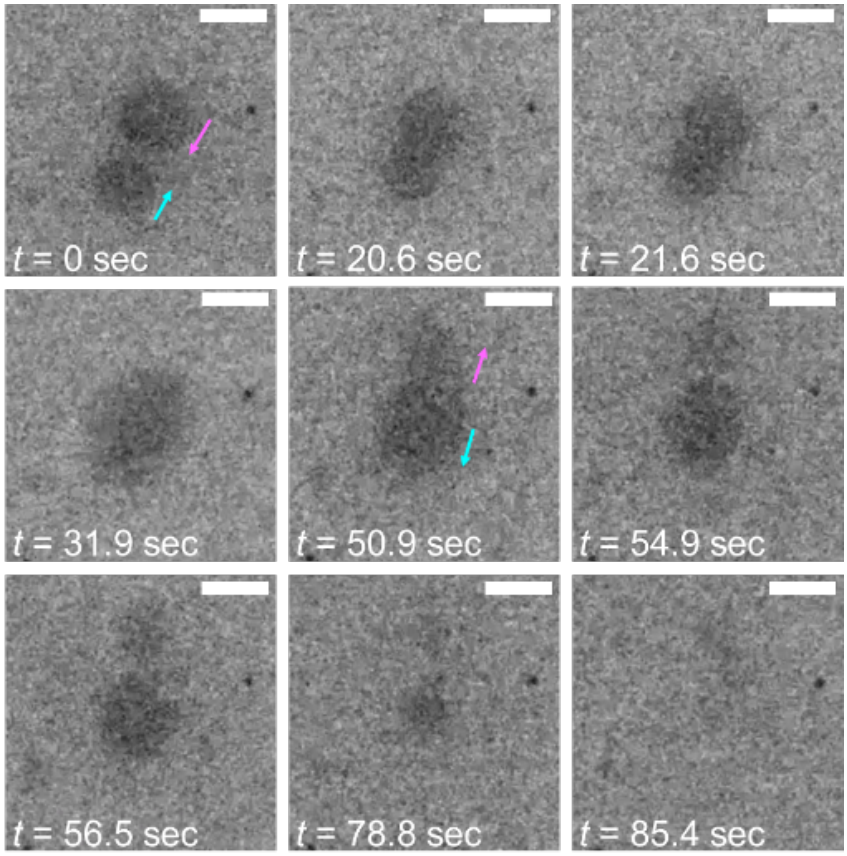


Fig. 4. Lifespan of zeptoliter liquid water droplets showing coalescence-budding-separation-evaporation dynamics. Scale bar is 200 nm.

3. Conclusions

In this work, we demonstrate that the nanoscale control of liquid water is feasible by theELC method. We believe that the well-defined dimension of liquid water is indispensable prerequisite to the total scattering measurements with TEM in various temperature and pressure conditions – that will provide a valuable insight into the molecular architecture of liquid water.

References

1. P. Ball.: Water — An enduring mystery. Nature 452(20), 291–292 (2008)
2. A. K. Soper: The radial distribution functions of water as derived from radiation total scattering experiments: Is there anything wecan say for sure? Int. Sch. Res. Notices Phys. Chem., 1–67(2013)
3. Ph. Wernet et al.: The structure of the first coordination shell inliquid water, Science 304, 995-999 (2004)
4. P. Gallo et al.: Water: a tale of two liquids. Chem. Rev. 116, 7463−7500 (2016)
5. J. A. Sellberg et al.: Ultrafast X-ray probing ofwater structure belowthehomogeneous ice nucleation temperature. Nature 510, 381-384 (2014)
6. A. K. Soper: Is water one liquid or two? J. Chem. Phys. 150, 234503 (2019)
7. N. d. Jonge et al.: Resolution and aberration correctionin liquid cell transmission electronmicroscopy. Nat. Rev. Mat. 4, 61–78 (2019)
8. J. M. Yuk et al.: High-resolution EM of colloidal nanocrystal growth usinggraphene liquid cells. Science 336, 61–64 (2012)
9. C. Mueller et al.: Nanofluidic cells with controlled pathlength and liquid flow for rapid, high-resolution in situ imaging with electrons. J. Phys. Chem. Lett. 4, 2339–2347 (2013)
10. M. N. Yesibolati et al.: Electron inelastic mean free path in water. Nanoscale 12, 20649–20656 (2020)

Magneto plasmonic responses in periodic structure

Jeong Woo Han[1,2]

[1]*Department of Physics Education and Center for Quantum Technologies, Chonnam National University, Gwangju 61186, South Korea*
[2]*Graduate School of Interdisciplinary Program for Photonic Engineering, Chonnam National University, Gwangju 61186, South Korea*

jwhan@chonnam.edu

Keywords: Terahertz, Graphene, Magneto plasmonics

1. Introduction

Strong optical anisotropy in the terahertz (THz) frequency range has been predicted for monolayer graphene under a static magnetic field. This property makes monolayer graphene a promising material for magnetically controlled polarization tuning of THz light through Faraday rotation (θ_F) [1]. However, the frequency at which the Faraday rotation angle is maximized strongly depends on the cyclotron resonance frequency, ω_c, which is determined by the applied magnetic field. This dependence makes it difficult to use monolayer graphene in practical devices for controlling the polarization of narrowband THz light. By applying a magnetic field to patterned graphene, magnetoplasmonic resonances can be induced at the magnetoplasmon frequency, ω_{mp}. Unlike ω_c, ω_{mp} lies near the plasmon resonance frequency, ω_p, of the patterned graphene and is less sensitive to the magnetic field [2,3]. Therefore, patterned graphene enables the frequency of maximum θ_F to be positioned near a desired ω_p.

In this study, we experimentally investigate Faraday rotation arising from magnetoplasmon excitation in graphene disks [4]. We observe a distinct difference in the magnetic-field dependence of θ_F between unpatterned graphene and graphene disks: notably, the sign of θ_F is opposite in the two systems. To understand this behavior, we simulate the experimental results by incorporating magnetoplasmonic effects and find good agreement with the measured data. These results suggest that magnetoplasmons provide a route to optimize θ_F at a target frequency. The quasi-freestanding bilayer graphene disks was fabricated by a hydrogen passivation technique on SiC substrate. The radius and the periodicity are 1.2 and 1.5 μm. From the transmission spectrum obtained from FT-IR, we confirm that the plasmon frequency ω_p, the scattering rate $\gamma=\Gamma/2\pi$, and the carrier density n are 3.57 THz, 9.31×10^{12} s^{-1}, and 7.8×10^{12} cm^{-2} respectively. Note that corresponding mobility μ_c and Fermi energy ε_F are about 3300 $cm^2/V\cdot s$ and 330 meV, which can be obtained from $\mu_c = ev_F/\sqrt{\pi n}\gamma\hbar$ and $\varepsilon_F=\sqrt{\pi n}\hbar v_F$[4]. Here, e and v_F are the elementary charge and Fermi-velocity (10^8 cm/s). To measure θ_F at ω_p, we employ a narrow band THz pulse (3.6 THz) generated by a free-electron laser (FELBE facility at Helmholtz-Zentrum Dresden-Rossendorf). The static magnetic field is changed from -6 T to 6 T, which applies to the perpendicular direction to the sample (See. Fig. 1(a)). All measurements are performed at a temperature of 10 K.

Optical conductivity tensor for unpatterned graphene is described as [4]:

$$\sigma_{xx}(B)=\sigma_{yy}(B)=q\frac{e^2|\varepsilon_F|}{\hbar^2\pi}\frac{i(\omega+i\Gamma)}{(\omega+i\Gamma)^2-(\omega_c(B))^2} \quad (1)$$

$$\sigma_{xy}(B)=-\sigma_{yx}(B)=q\frac{e^2|\varepsilon_F|}{\hbar^2\pi}\frac{\omega_c(B)}{(\omega+i\Gamma)^2-(\omega_c(B))^2} \quad (2)$$

For considering ω_c of bilayer graphene, we use the Landau level energy [6]. $q$is the scaling factor is employed, which takes into account the contribution of the graphene area. To account for modified plasmon frequency via applied magnetic fields, i.e., magnetoplasmonω_{mp}, ω_c in Eq. (1) and (2) is replaced by ω_{mp}[5]:

$$\omega_{mp}(B)=\sqrt{\frac{(\omega_c(B))^2}{4}+\omega_p^2}\pm\frac{|\omega_c(B)|}{2} \tag{3}$$

Note that + and - signs in Eq. (3) correspond to the direction of magnetic fields. Without magnetic fields, ω_{mp}isequal toω_p,indicating that Eq. (1) is regressed as the well-known optical conductivity for the plasmon resonance in patterned graphene [6]. Since σ_{xy}should play a role when only applying magnetic fields, we redefine the non-diagonal component of optical conductivity as $\sigma_{xy}[B]$-$\sigma_{xy}[0]$. Lastly, we compute $\theta_F$based on Eq. (1) – (3) via:

$$\theta_F=\frac{1}{2}\arg\left(\frac{t_{0xx}-it_{0xy}}{t_{0xx}+it_{0xy}}\right) \tag{4}$$

where t_{0xx} and t_{0xy} are the amplitudes for the transmitted plan waves of s and p polarizations. Detailed equation for t_{0xx} and $t_{0xy}$is defined in [7]. The substrate information is included in $t_{0xx}$and $t_{0xy}$and hence we use 12.25 as the dielectric constant for SiC substrate.

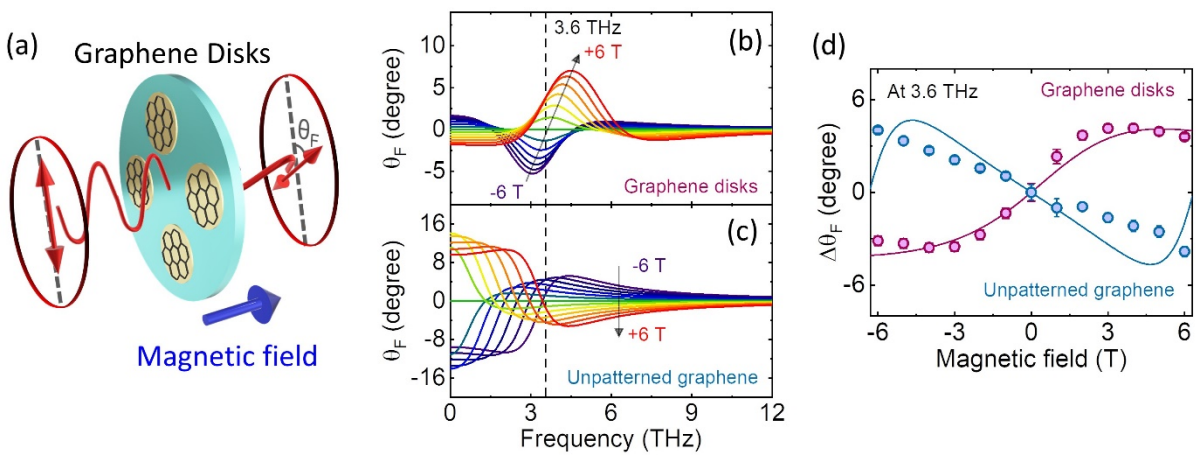


Fig. 1. (a) Experimental geometry. θ_F is the angle of Faraday rotation. (b)-(c) Simulated θ_F in graphene disks and unpatterned graphene. (d) $\Delta\theta_F$ at 3.6 THz. Line and symbols show the simulation result and experimental data, respectively. Reprinted from [4].

Figures 1(b) and 1(c) show simulated Faraday rotation angles, θ_F, for graphene disks and unpatterned graphene, respectively (reprinted from Ref. [4]). We used q = 0.9 for the graphene disks and q = 1 for the unpatterned graphene, indicating that the graphene disks were positioned accurately at the focal point of the incident THz beam. For the graphene disks, θ_F is strongly confined near the cyclotron frequency, ω_c. In contrast, θ_F varies dynamically in unpatterned graphene, in good agreement with the results reported in Ref.

[7]. Notably, in the unpatterned graphene, the sign of θ_F reverses near ω_c. This behavior originates from the sign change in the real part of σ_{xy} around ω_c. Figure 1(d) presents the magnetic-field dependence of $\Delta\theta_F$ at 3.6 THz, where $\Delta\theta_F(B) = \theta_F(B) - \theta_F(0)$. As shown, the simulations reproduce not only the overall profiles but also the amplitudes in both cases. The asymmetric behavior observed for the graphene disks near 0 T is attributed to the fact that ω_{mp} is not perfectly symmetric with respect to positive and negative magnetic fields. Together, the experimental and simulation results demonstrate that the frequency at which θ_F is optimized can be controlled by exploiting magnetoplasmons [4].

References

1. I. Crassee *et al*., "Giant Faraday rotation in single-and multilayer graphene," Nat. Phys. **7**, 48 (2011).
2. I. Crassee *et al*., "Intrinsic terahertz plasmons and magnetoplasmons in large scale monolayer graphene," Nano letters **12**, 2470 (2012).
3. J.-M. Poumirol *et al*., Electrically controlled terahertz magneto-optical phenomena in continuous and patterned graphene," Nat. Commun. **8**, 1 (2017).
4. J. W. Han et al., “Strong transient magnetic fields induced by THz-driven plasmons in graphene disks,” Nat. Commun. **14**, 7493 (2023).
5. M. L. Chin *et al*., "Observation of strong magneto plasmonic nonlinearity in bilayer graphene discs," JPhys photpnics **3**, 01LT01 (2021).
6. M. M. Jadidi *et al*., "Optical control of plasmonic hot carriers in graphene," ACS Photonics **6**, 302 (2019).
7. A. Ferreira *et al*., "Faraday effect in graphene enclosed in an optical cavity and the equation of motion method for the study of magneto-optical transport in solids," Phys. Rev. B **84**, 235410 (2011).

Hydrogen Lyman-α Emission in Astrophysics

Seok-Jun Chang

Max Planck Institute for Astrophysics, Garching, Germany
csj607@gmail.com

Abstract. Hydrogen is the most abundant element in the Universe, and its Lyman-$\boldsymbol{\alpha}$ (Ly$\boldsymbol{\alpha}$) transition is one of the most prominent emission lines observed across a wide range of astrophysical systems. Ly$\boldsymbol{\alpha}$ emission is produced through recombination and collisional excitation and plays an important role in the radiative cooling of hydrogen gas. At the same time, because Ly$\boldsymbol{\alpha}$ is a resonance line, photons can undergo numerous scatterings with neutral hydrogen before escaping from a system. As a result, the emergent Ly$\boldsymbol{\alpha}$ spectrum and spatial distribution carry rich information about the column density, temperature, geometry, and kinematics of the surrounding gas. Interpreting Ly$\boldsymbol{\alpha}$ observations thereforerequire an understanding of its radiative transfer and line-formation processes. In this article, I introduce the broad applications of Ly$\boldsymbol{\alpha}$ emission in astrophysics and review the basic physics of its interaction with neutral hydrogen. I then illustrate how Ly$\boldsymbol{\alpha}$ line profiles, spatial distributions, and polarization are formed through resonance scattering, with examples from three-dimensional Monte Carlo radiative-transfer simulations.

Keywords: Radiative transfer, Astrophysics, Line formation, Monte Carlo simulations

1. Introduction

Hydrogen Lyman-α (Lyα), corresponding to the $2p \rightarrow 1s$ transition at 1215.67 Å, is one of the most prominent emission lines in astrophysics. Lyα photons are produced mainly through recombination of ionized hydrogen and collisional excitation of neutral hydrogen, and the line is therefore closely associated with warm ionized and partially neutral gas [1].

A unique characteristic of Lyα is its resonant nature. Unlike many non-resonant emission lines, Lyα photons can undergo numerous scatterings with neutral hydrogen before escaping from an astrophysical system [1]. These interactions redistribute photons both in frequency and space, producing asymmetric or double-peaked spectra, extended Lyα halos, and polarized emission [2]. Consequently, the observed Lyα radiation depends not only on how the photons were originally produced but also on the column density, temperature, geometry, and kinematics of the neutral gas through which they propagate. This makes Lyα more difficult to interpret than many other emission lines, but at the same time makes it a particularly powerful probe of otherwise difficult-to-observe neutral hydrogen.

This same principle makes Lyα a valuable diagnostic across very different astrophysical settings, beginning with extragalactic astronomy. Although its rest-frame wavelength of 1215.67 Å lies in the ultraviolet, where Earth's atmosphere is opaque, the line is redshifted into the optical and near-infrared for distant sources, making it directly accessible to ground-based telescopes – a major practical reason it is one of the most widely used lines for studying galaxies in the early Universe. The interstellar and circumgalactic media around these galaxies can then be probed through Lyα radiative-transfer modeling [3,4]. Extended Lyα emission, sometimes called a Lyα blob, is commonly observed around galaxies and quasars on scales of tens to hundreds of kiloparsecs, powered by photoionization, collisional excitation, and the scattering of photons from the central source, and directly probing the surrounding cool, neutral gas [5,6,7]. Spectropolarimetric observations further show that the polarization of Lyα can independently constrain the geometry of this gas [8].

The importance of Lyα is not limited to galaxies. In young stellar systems, it can dominate the ultraviolet radiation field of accreting T Tauri stars, shaping the physical and chemical environment of their protoplanetary disks, and its profile encodes information about accretion flows, stellar winds, and surrounding neutral gas [9,10]. In the nearby TW Hya system, spatially resolved observations show that this scattering occurs both in the stellar wind and in neutral hydrogen within the disk itself [11].

2. Physics of Lyα Radiative Transfer

2.1. Line Profile and Cross Section

The Lyα absorption cross section as a function of frequency ν can be written as

$$\sigma_\nu = \frac{\pi e^2}{m_e c} f_\alpha \, \phi(\nu),$$

where $f_\alpha \approx 0.4162$ is the oscillator strength of the transition and $\phi(\nu)$ is a normalized line profile, $\int \phi(\nu)\, d\nu = 1$. Because Ly$\alpha$ photons are subject both to thermal (Doppler) motions of the absorbing atoms and to the finite radiative lifetime of the $2p$ state, $\phi(\nu)$ takes the form of a Voigt profile: a Gaussian core of width set by the Doppler parameter b, combined with Lorentzian damping wings set by the spontaneous decay rate $A_\alpha \approx 6.25 \times 10^8\ \mathrm{s}^{-1}$. In dimensionless frequency units $x \equiv (\nu - \nu_0)/\Delta\nu_D$ (with Doppler width $\Delta\nu_D = \nu_0\, b/c$), the ratio of the two widths defines the damping parameter, $a \simeq 4.7 \times 10^{-4}\,(T/10^4\ \mathrm{K})^{-1/2}$ for a purely thermal Doppler width.

The line-center optical depth for a neutral-hydrogen column density $N_{\mathrm{H\,I}}$ is $\tau_0 = N_{\mathrm{H\,I}}\,\sigma_{\nu_0}$. For the column densities typical of the systems discussed in Section 3 ($N_{\mathrm{H\,I}} \sim 10^{18}$–$10^{22}$ cm^{-2}), τ_0 can reach 10^5–10^8: an emitted Lyα photon is far more likely to be scattered than to escape directly, which is what makes its radiative transfer qualitatively different from that of non-resonant lines.

2.2. Frequency and Spatial Redistribution

At such large line-center optical depths, a photon emitted at $x = 0$ scatters many times before it can escape, performing a coupled random walk in frequency and position [1]. Two limiting behaviors govern how photons eventually escape: (i) in a static medium, repeated core scattering slowly diffuses photons into the damping wings, where the cross section is low enough to free-stream out; (ii) bulk gas motions further modify this frequency diffusion by changing the resonance condition along the photon path. Radial outflows or inflows can produce strong red-blue asymmetries, while rotation broadens the integrated profile and alters the peak separation depending on viewing angle [3,4].

The observational signature of this process is the characteristic double-peaked Lyα profile: photons escape preferentially in symmetric frequency ranges to the blue and red of line center, with a peak separation that grows with τ_0 (and hence $N_{\mathrm{H\,I}}$), while the relative strength and shift of the two peaks encode the presence of outflows, inflows, or rotation in the scattering medium. Because the same physical process also spreads photons spatially, it likewise produces extended, low-surface-brightness Lyα halos around the emitting source [1].

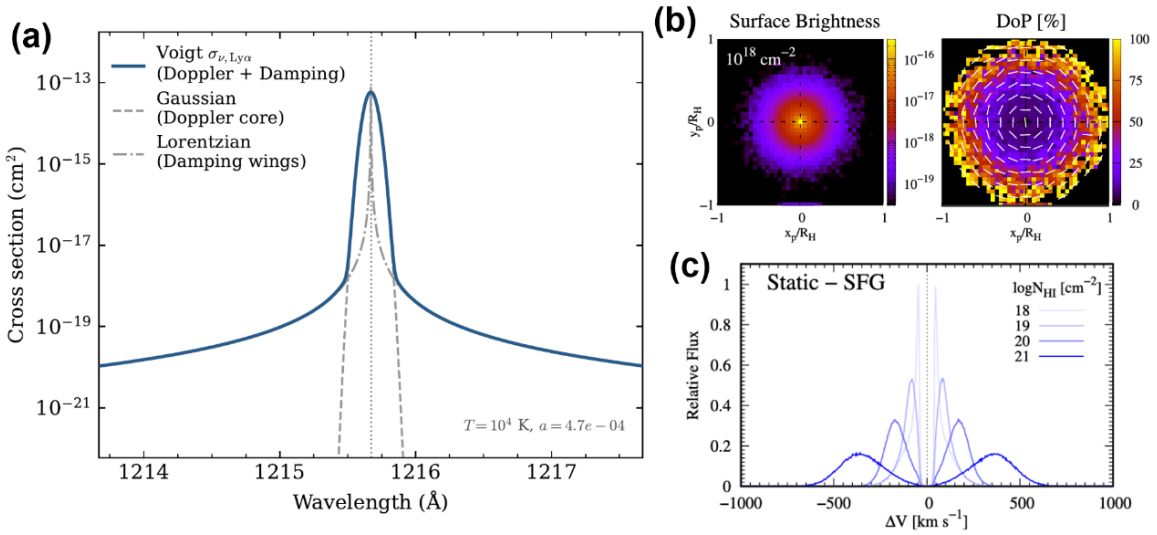


Fig. 1. Lyα line formation via resonant scattering: (a) the Voigt-profile scattering cross section, (b) surface-brightness and polarization maps of a spherical halo, and (c) line profiles for a static medium at increasing $N_{H\,I}$.

2.3 Polarization from Resonant Scattering

A further consequence of Lyα's resonant nature is that scattering off neutral hydrogen is not perfectly isotropic: like classical Rayleigh scattering, it depends on scattering angle, so radiation scattered through large angles emerges linearly polarized. Multiple scattering tends to average down this polarization but does not erase it, and the residual polarization provides information on the scattering geometry and photon propagation history that is complementary to that contained in the total intensity and line profile [2]. Figure 1 illustrates these effects with a 3D Monte Carlo radiative-transfer calculation: the scattering cross section itself, and the resulting surface-brightness, polarization, and line-profile structure it produces.

3. Lyα as a Probe Across Astrophysical Systems

The mechanisms described in Section 2 are generic to any optically thick neutral hydrogen medium, so the same handful of observables – line shape, spatial extent, and polarization – can probe systems on vastly different physical scales. We illustrate this with three examples: the circumgalactic medium, high-redshift star-forming galaxies, and protoplanetary disks around young stars, with observational examples of each shown in Figure 2.

Circumgalactic medium. The circumgalactic medium (CGM), the gas reservoir extending tens to hundreds of kiloparsecs around a galaxy, is one of the richest applications of Lyα radiative transfer. Because several distinct physical effects – column density, kinematics such as outflows or rotation, and geometry – can imprint similar signatures on the emergent spectrum, spatially unresolved observations alone are often insufficient to uniquely determine the underlying gas properties [3,5]. Spatially resolved spectroscopy, combined with radiative-transfer modeling of realistic, inhomogeneous gas distributions, helps break these degeneracies and more robustly connects the observed emission to the physical state of the CGM [4,6].

High-redshift star-forming galaxies. At the galaxy scale itself, only a few kiloparsecs across, polarization offers an observable that is particularly sensitive to scattering geometry and photon propagation history, complementing the information carried by the line profile alone. Spectropolarimetric observations of high-redshift star-forming galaxies can therefore constrain the geometry of their surrounding gas – for example the orientation of an outflow – in ways that complement, and help break degeneracies present in, purely spectral analyses [8]. Lyα can probe such galaxies not only object by object but also statistically across large populations: large-area narrowband imaging surveys can identify populations of extended Lyα nebulae in bulk, so that their number density, luminosity function, and enhancement in dense protocluster environments provide a statistical, population-level counterpart to the object-by-object view obtained from detailed spectral and polarimetric modeling [7].

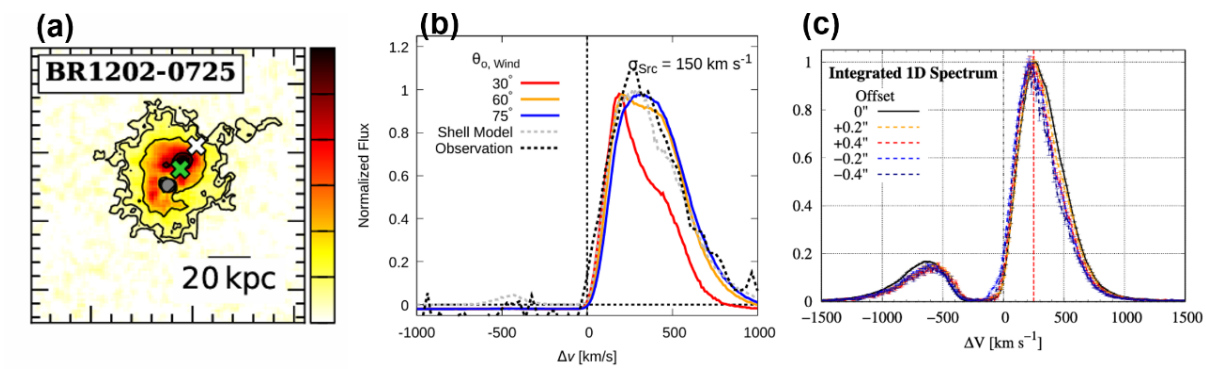


Fig. 2. Examples of Lyα observations and radiative-transfer modeling across diverse astrophysical environments: (a) the circumgalactic medium [5], (b) a high-redshift star-forming galaxy [8], and (c) a protoplanetary disk [11].

Young stellar objects and protoplanetary disks. On vastly smaller scales – protoplanetary disks extend only tens to a few hundred astronomical units (AU) from their host star, some six orders of magnitude below the CGM – Lyα radiation from accreting young stars both shapes and is shaped by the surrounding accretion flows, stellar and disk winds, and the protoplanetary disk itself. Modeling its scattered line profile, together with the fluorescent molecular emission it powers within the disk, provides diagnostics of wind and outflow kinematics that are otherwise difficult to obtain, and of the ultraviolet radiation field that regulates disk chemistry; for the nearby TW Hya system, spatially resolved observations probe the disk at offsets of order $\sim$10–25 AU from the star [9,10,11].

4. Conclusions

Hydrogen Lyα occupies a distinctive position among astrophysical emission lines: the same resonant scattering that makes it difficult to interpret is also what makes it uniquely informative. Because a photon typically scatters many times before escaping, its emergent spectrum, spatial distribution, and polarization jointly encode the column density, temperature, geometry, and kinematics of the neutral hydrogen it passed through. This single mechanism underlies applications on vastly different scales, from the circumgalactic medium and star-forming galaxies in the early Universe to accretion and outflows around young stars and their protoplanetary disks. Because several physical properties can imprint similar signatures on the line alone, robust interpretation increasingly relies on combining multiple observables with three-dimensional Monte Carlo radiative-transfer modeling – making Lyα likely to remain one of the most versatile tools for studying neutral hydrogen across astrophysics.

References

1. Chang, S.-J., Yang, Y., Seon, K.-I., Zabludoff, A., Lee, H.-W.: Radiative transfer in Lyα nebulae. I. Modeling a continuous or clumpy spherical halo with a central source. The Astrophysical Journal **945**, 100 (2023).
2. Chang, S.-J., Lee, H.-W., Yang, Y.: Polarization of Rayleigh scattered Lyα in active galactic nuclei. Monthly Notices of the Royal Astronomical Society **464**, 5018-5027 (2017).
3. Kym, H.-G., Chang, S.-J., Gronke, M., Seon, K.-I.: The impact of circumgalactic rotation on Lyα radiative transfer. arXiv:2609.08396 (2026). Submitted to MNRAS.
4. Chang, S.-J., Gronke, M.: Probing cold gas with Mg II and Lyα radiative transfer. Monthly Notices of the Royal Astronomical Society **532**, 3526-3555 (2024).
5. González Lobos, V. et al.: Circumgalactic Lyα emission around submillimeter-bright galaxies with different quasar contributions. Astronomy & Astrophysics **679**, A41 (2023).
6. Almada Monter, S., Gronke, M., Chang, S.-J.: Lyman-α escape through anisotropic media. Monthly Notices of the Royal Astronomical Society **547**, 1-17 (2026).
7. Moon, B. et al.: ODIN: A new Lyα blob selection method, sample, and statistical analysis at $z \sim 3.1$. The Astrophysical Journal **998**, 345 (2026).
8. Bolamperti, A. et al.: Constraining the geometry of the gas surrounding a typical galaxy at $z = 3.4$ with Lyα polarization. Astronomy & Astrophysics **695**, A119 (2025).
9. Arulanantham, N. et al.: Lyα scattering models trace accretion and outflow kinematics in T Tauri systems. The Astrophysical Journal **944**, 185 (2023).
10. Kalscheur, M. et al.: Evidence of a disk wind origin for fluorescent H_2 in classical T Tauri stars. The Astronomical Journal **169**, 240 (2025).
11. Chang, S.-J., Arulanantham, N., Gronke, M., Herczeg, G.J., Bergin, E.A.: First detection and modelling of spatially resolved Lyα in TW Hya. Monthly Notices of the Royal Astronomical Society **529**, 2656-2675 (2024).

Life Science Division

Aims

The Life Science Division invites voluntary contributions from VeKNI members that reflect the broad and rapidly evolving research landscape across the life sciences. Members are encouraged to share recent scientific advances, summaries of ongoing research projects, emerging technology trends, translational applications, industrial case studies, or perspectives on future challenges in biological and biomedical research.

Scope

- **Molecular & Cellular Life Sciences:** Molecular biology, cell biology, genetics, developmental biology, physiology, neuroscience, immunology, and disease mechanisms

- **Biomedical & Translational Research:** Human health and disease, diagnostics, therapeutics, regenerative medicine, precision medicine, drug discovery, and clinical or preclinical research

- **Omics & Computational Biology:** Genomics, transcriptomics, proteomics, metabolomics, epigenomics, single-cell and spatial omics, bioinformatics, systems biology, and AI-driven biological data analysis

- **Biotechnology & Bioengineering:** Synthetic biology, genome engineering, organoids, tissue engineering, bioprocessing, biosensors, biomolecular engineering, and industrial biotechnology

- **Microbiology & Host–Environment Interactions:** Microbial biology, microbiome research, infectious diseases, host–pathogen interactions, antimicrobial resistance, and environmental microbiology

- **Ecology, Evolution & Environmental Life Sciences:** Biodiversity, evolution, marine and fisheries science, plant and animal biology, ecosystem health, climate adaptation, and environmental sustainability in living systems

Visualizing Proteins at Work: Time-Resolved Serial Crystallography and Sample Support at the European XFEL

Huijong Han

European XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany
huijong.han@xfel.eu

Abstract. X-ray free-electron lasers have made it possible to record protein structures while they react. By combining femtosecond pulses with serial measurement, time-resolved serial crystallography captures reaction intermediates at room temperature and essentially free of radiation damage. This article outlines the principles of the method, representative experiments at the European XFEL, and the role of the XBI BioLab in making them feasible.

Keywords: Serial femtosecond crystallography, time-resolved structural biology, X-ray free-electron laser, European XFEL, XBI

Introduction

Since the first protein structures were solved by X-ray crystallography in the late 1950s,[1,2] it has produced most entries in the Protein Data Bank (PDB) and has supportedmost of modern structure-based drug design. This success rests on two compromises. To limit radiation damage, data are conventionally collected near 100 K,[3] which suppresses functionally important conformational heterogeneity. More importantly for this article, a conventional dataset averages over an exposure far longer than almost any catalytic or signaling event, so a cryo-cooled structure is a still photograph of a machine that only makes sense in motion. Time-resolved Laue diffraction showed three decades ago that intermediates can be visualized,[4] but it required large crystals that survive repeated exposure.

The decisive step came from the recognition that a sufficiently short X-ray pulse allows diffraction to outrun the destruction it causes. A pulse of a few tens of femtoseconds terminates before ionization-induced atomic displacements develop, so the pattern reflects an essentially intact structure even though the sample is then vaporized.[5] This principle was proved at the Linac Coherent Light Source (LCLS),[6] and high-resolution structures from micrometer-sized crystals followed.[7] The resulting method, serial femtosecond crystallography (SFX), replaces the rotation of one large crystal with single still patterns from many small ones in random orientations. Because every crystal is measured once and then replaced, dose becomes irrelevant, room-temperature collection becomes routine, and a reaction can be triggered in each fresh crystal at a defined moment before it is probed. Repeating this at different delays yields a time course, first demonstrated for photoactive yellow protein (PYP).[8]

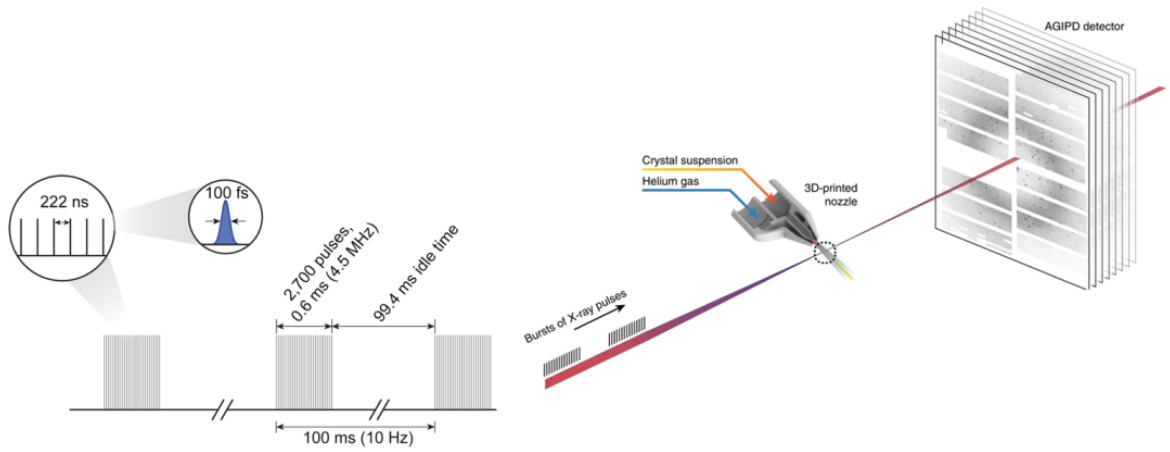


Fig. 1. Schematic SFX experiment with EuXFEL pulse structure. (right side image is from Wiedorn*et al.* (2018)[9])

European XFEL

A time-resolved experiment couples four elements. Homogeneous microcrystals, a delivery system that renews the sample between pulses, a trigger, and a detector and analysis chain able to handle the data rate. Hits must be identified among a large majority of blank frames and indexed without prior knowledge of orientation, with tens of thousands of patterns typically required per time point.[10] The result is extracted as difference electron density maps between an activated and a reference dataset.

The European XFEL is the first hard X-ray FEL capable of megahertz repetition rates, delivering pulses in trains at 10 Hz with MHz spacing within each train.[11] Early experiments established that serial crystallography works under these conditions and that structures from successive pulses of a train are indistinguishable, including for membrane-protein complexes as large as Photosystem I.[9,12–14] What the repetition rate contributes is not better time resolution, which is set by the pump and probe pulse durations, but statistics. More patterns per unit beamtime means smaller detectable difference signals and more time points per experiment, which for weakly populated intermediates decides whether a map is interpretable at all. Most SFX experiments performed at the SPB/SFX instrument[15], with pump-probe operation requiring optical lasers synchronized to the pulse train.[16]

Reaction Initiation and Representative Experiments

Light-triggered systems. Where the protein carries a chromophore, the reaction is initiated by an optical pulse and the pump-probe delay defines the time point. Delivery is typically by gas dynamic virtual nozzle,[17] which produces a jet fast enough for MHz operation, or by viscous extrusion for membrane proteins in lipidic cubic phase.[18] The first molecular movie at the European XFEL followed the PYP photocycle, driven by trans to cis isomerization of the para-coumaric acid chromophore.[19] Working in the picosecond range, Pandey and co-workers reproduced results from lower-repetition-rate facilities, reached a previously unmapped time range, and established that the pump laser and the pulse train can be interleaved without exciting a crystal twice.

Light-independent systems. Most enzymes are not photoactive. Mix-and-inject serial crystallography combines the crystal slurry with substrate shortly before the interaction point, so the delay is set by transit time.[20] Time resolution is bounded by substrate diffusion into the crystal, which favors small crystals, and spans microseconds to seconds. The approach was first demonstrated on β-lactamase from Mycobacterium tuberculosis reacting with ceftriaxone[21] and later transferred to the European XFEL, where diffusion and binding in the same enzyme were separated on millisecond timescales.[22] Sample consumption remains the principal obstacle, since continuous injection discards most material between pulses. A segmented droplet scheme applied to the human flavoenzyme NQO1 recently cut consumption by roughly 97 percent while preserving data quality.[23]

The XBI BioLab

XFEL experiments are far more likely to fail because of the sample than because of the X-ray source. Crystal size and density, sample aggregation, and buffer viscosity all determine whether a sample jet operates stably and whether sufficient diffraction hits can be collected. In addition, biological samples rarely remain in optimal condition after prolonged transport.

To address these challenges, the XFEL Biology Infrastructure (XBI) was established as a User Consortium, with funding and personnel provided by partners in Finland, Germany, the Slovak Republic, Sweden, and the USA.[24] The BioLab, which opened in 2016, is located in the European XFEL headquarters directly above the experimental instruments. Covering the entire workflow from gene to sample injection, it provides facilities for cell culture, protein purification, biophysical characterization, crystallization of both soluble proteins and samples in lipidic cubic phase, as well as an injection test chamber.

Several of these capabilities are directly relevant to time-resolved experiments. Reproducible crystal size determines the achievable time resolution in mixing experiments and promotes uniform photoactivation in pump–probe studies. Spectroscopic characterization guides the selection of pump wavelength and fluence, ensuring excitation of a sufficient fraction of the sample while minimizing multiphoton artifacts. Offline injection testing validates sample delivery before beamtime rather than during valuable XFEL access. As an example, NQO1 crystals were grown on site for both beamtime campaigns, with the crystal size adjusted between experiments to match the injector requirements.[23] This illustrates the broader advantage of on-site sample preparation, which enables rapid optimization when initial experimental conditions prove suboptimal.

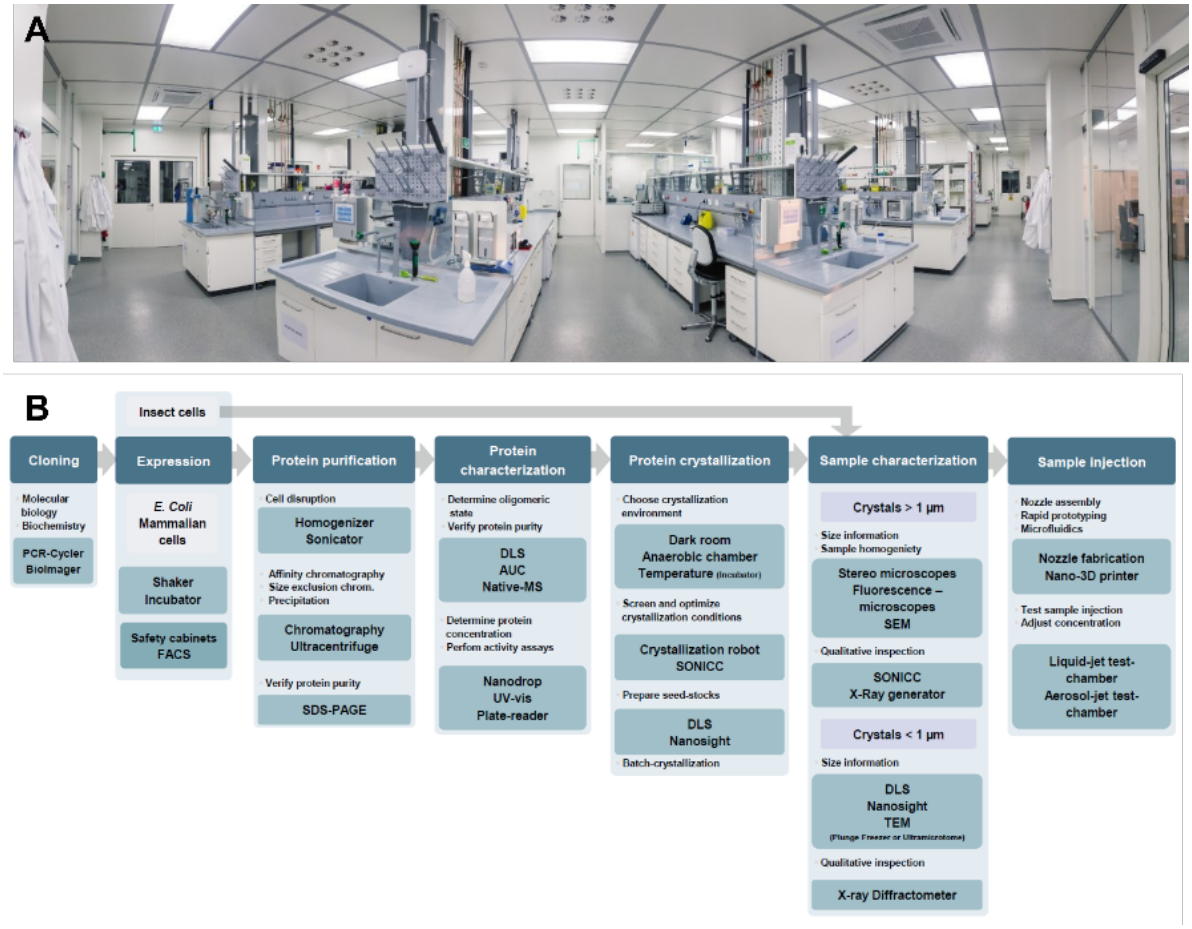


Fig. 2. A) A photo of XBI BioLab. B) A workflow scheme for SFX sample preparation and characterization at XBI BioLab.

Outlook

For many targets, the quantity of microcrystals required remains the primary limiting factor, and further reductions in sample consumption are likely to broaden the accessible range more than any increase in source performance. Data volumes now demand near-real-time feedback and automated pipelines, and difference maps at low activation are easily over-read, so cross-validation against spectroscopy and simulation remains essential. With synchrotrons covering timescales from microseconds to seconds and cryo-EM providing complementary structural information, XFEL-based studiesare best seen as one component of an integrated structural biology. Their unique capability, probing structural dynamics from femtosecond to microsecond at room temperature without radiation damage, is now expanding from model systems to biologically and medically relevant targets, including drug targets and signaling mechanisms. Realizing this potential depends as much on advances in sample preparation and delivery as on continued improvements in the X-ray source, a challenge that the XBI BioLab was established to address.

References

1.Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181** (1958).

2.Perutz, M. F. *et al.*Structure of Hæmoglobin: A three-dimensional fourier synthesis at 5.5-. resolution, obtained by X-ray analysis. *Nature* **185** (1960).

3.Garman, E. F. Radiation damage in macromolecular crystallography: What is it and why should we care? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66** (2010).

4.Šrajer, V. *et al.* Photolysis of the carbon monoxide complex of myoglobin: Nanosecond time-resolved crystallography. *Science* **274** (1996).

5.Neutze, R., Wouts, R., Van Der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses.

Nature **406** (2000).
6.Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470** (2011).
7.Boutet, S. *et al.* High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science* **337** (2012).
8.Tenboer, J. *et al.* Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346,** 1242–1246 (2014).
9.Wiedorn, M. O. *et al.*Megahertz serial crystallography. *Nat. Commun.* **9** (2018).
10.White, T. A. *et al.* CrystFEL: A software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **45** (2012).
11.Decking, W. *et al.* A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator. *Nat. Photonics* **14** (2020).
12.Grünbein, M. L. *et al.*Megahertz data collection from protein microcrystals at an X-ray free-electron laser. *Nat. Commun.* **9** (2018).
13.Yefanov, O. *et al.* Evaluation of serial crystallographic structure determination within megahertz pulse trains. *Struct. Dyn.* **6** (2019).
14.Gisriel, C. *et al.* Membrane protein megahertz crystallography at the European XFEL. *Nat. Commun.* **10** (2019).
15.Mancuso, A. P. *et al.* The single particles, clusters and biomolecules and serial femtosecond crystallography instrument of the european XFEL: Initial installation. *J. Synchrotron Radiat.* **26** (2019).
16.Koliyadu, J. C. P. *et al.*Pump-probe capabilities at the SPB/SFX instrument of the European XFEL. *J. Synchrotron Radiat.* **29** (2022).
17.DePonte, D. P. *et al.*Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D. Appl. Phys.* **41** (2008).
18.Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 3309 (2014).
19.Pandey, S. *et al.* Time-resolved serial femtosecond crystallography at the European XFEL. *Nat. Methods* **17** (2020).
20.Schmidt, M. Mix and inject: Reaction initiation by diffusion for time-resolved macromolecular crystallography. *Adv. Condens. Matter Phys.* **2013**, 167276 (2013).
21.Kupitz, C. & et al. Structural enzymology using X-ray free electron lasers Open Access . *Struct. Dyn.* **4** (2017).
22.Pandey, S. *et al.* Observation of substrate diffusion and ligand binding in enzyme crystals using high-repetition-rate mix-and-inject serial crystallography. *IUCrJ* **8**, 878–895 (2021).
23.Doppler, D. *et al.* Minimized sample consumption for time-resolved serial crystallography applied to the redox cycle of human NQO1. *Commun. Chem.* **9** (2026).
24.Han, H. *et al.* The XBI BioLab for life science experiments at the European XFEL. *J. Appl. Crystallogr.* **54**, 7–21 (2021).

Convergence of Single-Cell Omics and Spatial Omics

Hyun-Woo Jeong[1]

[1] Single-cell Multiomics Laboratory, Max Planck Institute for Molecular Biomedicine, Röntgenstraße 20, 48149Münster, Germany
Hyun-woo.jeong@mpi-muenster.mpg.de

Abstract. Single-cell omics and spatial omics are moving from cataloguing cell types toward measuring regulatory state, molecular mechanism, perturbation response, and tissue context in the same analytical frame. The latest research trend is therefore not one platform replacing another, but convergence: higher modality per cell, finer spatial resolution in tissue, and larger computational models that align dissociated and in situ measurements. This short review summarizes the technology shifts that now define the field, including single-cell multiome profiling, Perturb-seq, spatial multi-omics, subcellular spatial transcriptomics, and foundation-model analysis.

Keywords: single-cell multi-omics; spatial transcriptomics; spatial multi-omics; foundation models

1. Introduction: from cell lists to spatial mechanisms

Single-cell RNA sequencing established cell-state atlases, but current omics development is increasingly driven by two limitations of RNA-only cell dissociation: cells lose tissue coordinates, and one molecular layer rarely explains mechanism by itself. New platforms address these gaps by measuring chromatin accessibility, surface protein, lineage or perturbation barcodes, and spatial position alongside transcriptomes. In parallel, computational methods have shifted from clustering isolated data sets toward integration, transfer learning, and predictive representations. The practical consequence is a field organized around three questions: what state is the cell in, what regulatory program caused it, and where does it act in tissue?

2. Single-cell omics: multimodal and perturbational readouts

The dominant technical direction in single-cell omics is co-assay rather than deeper measurement of a single layer. Single-cell RNA and chromatin accessibility co-profiling, exemplified by SNARE-seq, made it possible to link gene expression states to cis-regulatory landscapes in the same cell [1]. Protein-and-RNA strategies such as CITE-seq added antibody-derived tags to transcriptomic profiles, improving immune phenotyping and resolving cases where surface markers and mRNA abundance diverge [2]. These concepts have matured into commercial multiome workflows and increasingly into experiment designs that pair genotype, epigenome, transcriptome, surface phenotype, and cell hashing for sample multiplexing.

A second trend is causal perturbation at atlas scale. Genome-scale Perturb-seq maps guide-RNA identity to single-cell transcriptional phenotypes, allowing researchers to read pathway state, buffering, and genetic interaction structure rather than only growth or marker outcomes [3]. This turns single-cell profiling into a functional genomics assay: the same experiment can identify cell-state regulators, infer gene programs, and nominate therapeutic targets. The main challenge is no longer only capture efficiency, but experimental design under combinatorial perturbations, batch structure, and cell-state heterogeneity.

Computational integration has become a necessary part of the measurement stack. MultiVI showed how a probabilistic latent model can align single-cell multiomic data even when RNA and ATAC are not always observed in the same cell [4]. Benchmarking studies now emphasize that the best method depends on task, modality balance, and whether prediction, batch correction, or biological conservation is the primary objective [5]. More recently, foundation models such as scGPT reframed single-cell data as a pretraining corpus for transfer to perturbation prediction, cell-type annotation, and cross-study representation learning [6]. These models are promising, but they also make negative controls, external validation, and transparent evaluation more important.

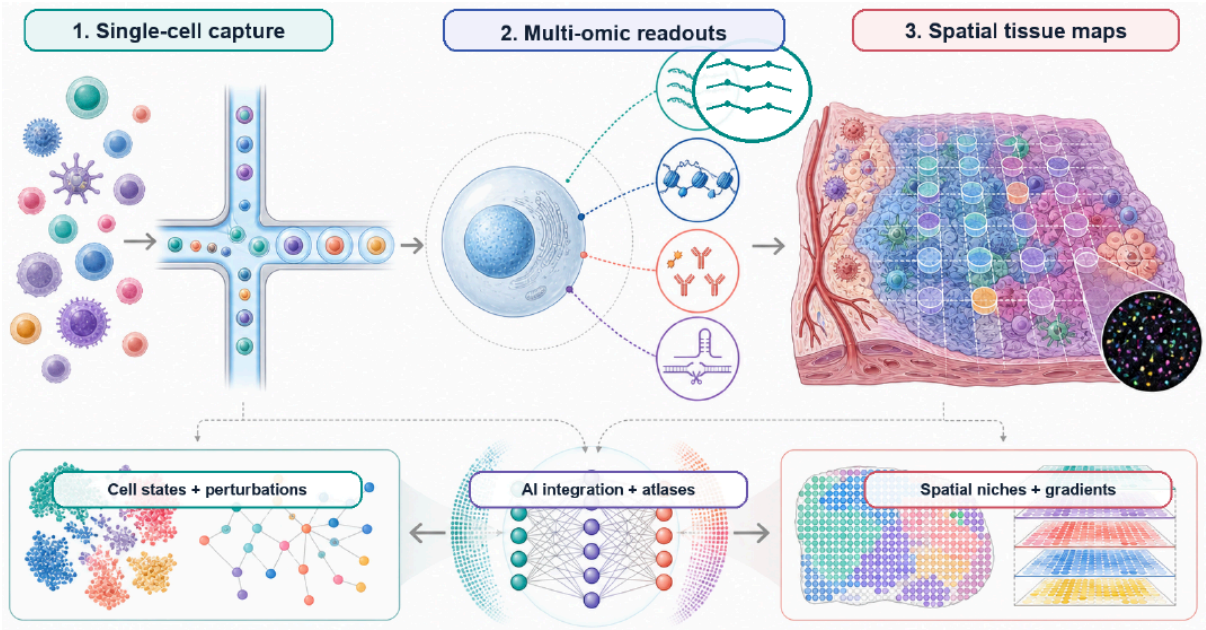


Fig. 1. The schematic summarizes current technology development: dissociated single cells enable paired molecular readouts and perturbation maps; intact tissue assays add coordinates, neighborhoods, and subcellularcontext; computational models integrate both streams into reusable atlases and predictive representations. Base image generated with imageGen; labels overlaid deterministically.

3. Spatial omics: resolution, modality, and tissue architecture

Spatial omics has expanded along two axes: more molecules per coordinate and smaller coordinates per molecule. Sequencing-based spatial methods began with array barcoding and have advanced to higher-density capture surfaces. DBiT-seq introduced deterministic barcoding in tissue, enabling high-spatial-resolution co-measurement of RNA and protein and providing a route toward spatial multi-omics rather than transcript-only maps [7]. Spatial-CITE-seq further demonstrated high-plex protein and whole-transcriptome co-mapping at cellular resolution [8], directly connecting tissue niches to immune and stromal phenotypes.

At organ and developmental scale, DNA nanoball-patterned arrays in Stereo-seq generated a spatiotemporal mouse organogenesis atlas, illustrating how spatial transcriptomics can capture gradients, morphogenesis, and cell-state transitions across whole tissues [9]. Imaging-based methods, especially MERFISH, push resolution to subcellular single-molecule localization and have enabled a molecularly defined whole mouse brain atlas [10]. The newest platform benchmarking work across human tumors highlights the practical trade-offs among Visium HD, Xenium, CosMx, and other high-throughput systems: gene panel size, sensitivity, segmentation, field of view, FFPE compatibility, subcellular localization, cost, and computational burden all shape the right choice for a study [11].

4. Research trends and near-term outlook

Three research trends now stand out. First, spatial context is becoming a mechanism-discovery layer. In cancer, fibrosis, neurobiology, and immunology, the target unit is often a niche: a tumor-immune boundary, vascular-fibroblast module, synaptic microenvironment, or inflammatory lesion compartment. Second, single-cell reference atlases are increasingly used as priors for spatial deconvolution and cell-type localization, while spatial data feedback by revealing neighborhoods and ligand-receptor interfaces that dissociated data cannot see. Third, AI models are moving from annotation toward perturbation and virtual spatial profiling, but their credibility depends on prospective validation and careful separation of biological signal from batch, tissue, and platform confounders.

For the next several years, the most productive studies will likely combine targeted depth with broad context: a single-cell multiome or Perturb-seq experiment to identify regulatory candidates, followed by spatial transcriptomic or spatial proteogenomic validation in intact tissue. Methodologically, the field needs better standards for tissue processing, cell segmentation, cross-platform calibration, spatial false-discovery control, and privacy-preserving sharing of human tissue atlases. Conceptually, the field is moving from 'which cell types are present?' to 'which cell states are interacting, where, under which regulatory programs, and with what causal consequences?' That shift is the central reason single-cell omics and spatial omics should now be treated as one connected research ecosystem.

5. Key take-home messages

- The center of gravity is shifting from RNA-only cell atlases to multimodal, perturbational, and spatially anchored measurements.
- Spatial omics adds the missing tissue grammar: neighborhoods, boundaries, gradients, and subcellular localization.
- Foundation models are useful when they improve transfer and prediction, but prospective validation remains essential.

References

1. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat Biotechnol. 2019;37(12):1452-1457. PMID: 31611697. DOI: 10.1038/s41587-019-0290-0.
2. Stoeckius M, Hafemeister C, Stephenson W, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017;14(9):865-868. PMID: 28759029. DOI: 10.1038/nmeth.4380.
3. Replogle JM, Saunders RA, Pogson AN, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. Cell. 2022;185(14):2559-2575.e28. PMID: 35688146. DOI: 10.1016/j.cell.2022.05.013.
4. Ashuach T, Gabitto MI, Koodli RV, et al. MultiVI: deep generative model for the integration of multimodal data. Nat Methods. 2023;20(8):1222-1231. PMID: 37386189. DOI: 10.1038/s41592-023-01909-9.
5. Hu Y, Wan S, Luo Y, et al. Benchmarking algorithms for single-cell multi-omics prediction and integration. Nat Methods. 2024;21(11):2182-2194. PMID: 39322753. DOI: 10.1038/s41592-024-02429-w.
6. Cui H, Wang C, Maan H, et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods. 2024;21(8):1470-1480. PMID: 38409223. DOI: 10.1038/s41592-024-02201-0.
7. Liu Y, Yang M, Deng Y, et al. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. Cell. 2020;183(6):1665-1681.e18. PMID: 33188776. DOI: 10.1016/j.cell.2020.10.026.
8. Liu Y, DiStasio M, Su G, et al. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. Nat Biotechnol. 2023;41(10):1405-1409. PMID: 36823353. DOI: 10.1038/s41587-023-01676-0.
9. Chen A, Liao S, Cheng M, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell. 2022;185(10):1777-1792.e21. PMID: 35512705. DOI: 10.1016/j.cell.2022.04.003.
10. Zhang M, Pan X, Jung W, et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. Nature. 2023;624(7991):343-354. PMID: 38092912. DOI: 10.1038/s41586-023-06808-9.
11. Ren P, Zhang R, Wang Y, et al. Systematic benchmarking of high-throughput subcellular spatial transcriptomics platforms across human tumors. Nat Commun. 2025;16(1):9232. PMID: 41107232. DOI: 10.1038/s41467-025-64292-3.

Fates and Generation of Supernumerary Centrioles

Byungho Shin[1]

[1]Max Planck Institute for Molecular Biomedicine, Röntgenstraße 20, 48149 Münster-Gievenbeck, Germany

Byungho.shin@mpi-muenster.mpg.de

Abstract. Centrioles form the core of centrosomes and contribute to spindle assembly and ciliogenesis. Because centriole duplication is normally restricted to once per cell cycle, numerical abnormalities usually reflect a failure to coordinate centrosome biogenesis with cell-cycle progression. Excess centrioles are common in cancer, but their effects vary: some promote mitotic delay and chromosome missegregation, whereas others are structurally incomplete and functionally weak. Here, four major routes to centriole amplification—whole-genome doubling, PLK4 hyperactivity, centriole over-elongation, and premature centriole separation after PCM disruption—are summarized together with the principal cellular responses to these abnormalities.

Keywords: Centrosome, Supernumerary centrioles, Cancer cells.

1. Centriole number is coupled to the cell cycle

A proliferating animal cell typically contains two centrosomes, each organized around a centriole pair embedded in pericentriolar material (PCM). A procentriole forms next to each mother centriole around the G1/S transition, and the newly formed daughter remains engaged with the mother until mitotic exit. After disengagement, the daughter gradually acquires the capacity to recruit PCM, organize microtubules, and support another round of centriole assembly [1,2,3].

PLK4 is a central kinase in this process. At the G1/S transition, PLK4 cooperates with STIL to recruit SAS6 and build the cartwheel that defines the ninefold architecture of a new centriole [4,5]. PLK4 activity is therefore tightly controlled, while mother-daughter engagement provides an additional local block to reduplication [6,7]. Failure of either regulatory layer can generate supernumerary centrioles.

2. Major routes to supernumerary centriole formation

Whole-genome doubling and failed cytokinesis

Cytokinesis failure, cell fusion, endoreplication, or mitotic slippage can generate tetraploid cells that inherit twice the normal chromosome and centrosome complement [8]. Whole-genome doubling precedes aneuploid evolution in a substantial fraction of tumors, yet the accompanying extra centrosomes are often not stably maintained [9,10]. During subsequent divisions, cells carrying excess centrosomes divide inefficiently and may be eliminated, whereas clones that reduce centrosome number can preferentially expand [11,12].

The p53 pathway is an important barrier to the persistence of centrosome amplification. Extra centrosomes can prolong mitosis and activate stress-responsive pathways that stabilize p53, including Hippo-LATS2 signaling and the centrosome-associated PIDDosome [13,14,15]. Loss of this surveillance environment can therefore favor continued proliferation of cells with abnormal centrosome numbers.

Excessive PLK4 activity

Elevating PLK4 activity can generate multiple procentrioles around a parental centriole without first changing ploidy [16,17]. The resulting phenotype demonstrates that centriole amplification alone is sufficient to disturb spindle organization, although its biological consequences depend strongly on tissue context and genetic background.

In several developing tissues, PLK4-driven centrosome amplification causes aneuploidy, apoptosis, progenitor loss, or microcephaly rather than spontaneous tumor formation [18,19,20]. By contrast, centrosome amplification can accelerate tumorigenesis when p53 function is compromised, and chronic or transient PLK4 elevation has been shown to promote tumor formation in mammalian models [21,22].

Centriole over-elongation

Numerical amplification may also arise from structural abnormalities. Centrioles become excessively long when elongation factors such as CPAP are overproduced, when PLK1- or PLK2-dependent regulation is perturbed, or when cells remain in G2 or mitosis for an extended period [23,24,25]. Severely elongated centrioles can fragment or support ectopic procentriole formation along their walls, converting a length-control defect into a numerical abnormality [26,27].

Premature centriole separation after PCM disruption

During early mitosis, mother and daughter centrioles are disengaged but remain physically associated within a shared PCM scaffold [28,29]. Pericentrin cleavage and PCM remodeling normally occur in a controlled manner near mitotic exit [30,31]. If PCM integrity is lost too early, the centriole pair can separate prematurely. Because mitotic cells retain PLK4 activity, these liberated centrioles may initiate additional assembly during M phase [32,33].

Importantly, M-phase-generated centrioles are not necessarily equivalent to normal S-phase products. In TP53;PCNT;CEP215-deficient cells, such centrioles show defective daughter-to-mother conversion, poor PCM recruitment, reduced microtubule-organizing capacity, and limited ability to duplicate in the next cycle [33]. Thus, centriole number alone does not indicate how many fully functional centrosomes are present.

3. Cellular fates of excess centrioles

Selection against amplified cells

Cells with supernumerary centrioles often progress more slowly through interphase and mitosis, creating a measurable fitness disadvantage [34]. Multipolar spindle formation can cause arrest, cell death, or severe chromosome imbalance, leading to selective loss of highly amplified cells from a population [35]. Accordingly, centrosome-number homeostasis can emerge through population-level selection rather than through active correction in every individual cell.

Centrosome clustering

The best-characterized adaptation is centrosome clustering. Cells can gather several centrosomes into two functional spindle poles, producing a pseudobipolar spindle that supports continued division [36]. Clustering depends on force-generating proteins such as dynein and KIFC1/HSET, as well as cortical and kinetochore-associated factors that influence spindle organization [37,38,39]. Cancer cells often perform this process more efficiently than non-transformed cells [11].

Clustering nevertheless carries a cost. Transient multipolar spindle geometry can produce merotelic kinetochore-microtubule attachments and chromosome missegregation [11]. Conversely, asymmetric clustering can distribute a more normal centrosome complement to one daughter cell and thereby reduce centrosome number during the early evolution of tetraploid cells [40].

Functional inactivation

Another way to tolerate excess centrioles is to render a subset functionally inactive. Defective centrioles generated outside the normal S-phase program may fail to recruit mature PCM, organize microtubules, or duplicate [33]. Cells may therefore contain more centriole-like structures than active spindle-organizing centrosomes, which helps explain why some highly amplified cells still assemble an apparently bipolar spindle.

4. Biological significance and open questions

Supernumerary centrioles are heterogeneous in both structure and activity. Fully competent extra centrosomes can severely perturb spindle geometry, whereas immature or damaged centrioles may have limited immediate effects [33,27]. Future studies should therefore assess centriole architecture, duplication competence, and microtubule-organizing activity rather than centriole number alone.

Centrosome amplification may also influence tumor behavior outside mitosis. Increased microtubule nucleation can activate Rac1-dependent actin remodeling and promote invasion, while structural centrosome abnormalities can disrupt epithelial organization in three-dimensional culture [41,42]. These effects may

provide selective benefits that counterbalance the substantial mitotic costs of maintaining extra centrosomes.

Finally, the mechanisms that remove defective centrioles remain incompletely understood. Population-level selection is clearly important, but intracellular degradation pathways may also contribute. Autophagy controls centrosomal components such as CEP63, suggesting that altered centrosome-phagy could affect both the generation and persistence of abnormal centrioles [43,44]. A complete model of centrosome amplification will therefore need to integrate centriole production, structural quality, checkpoint responses, and clonal selection.

References

1. Wang WJ, Soni RK, Uryu K, Tsou MF. The conversion of centrioles to centrosomes: essential coupling of duplication with segregation. J Cell Biol. 2011;193:727-739.
2. Fu J, Lipinszki Z, Rangone H, Min M, Mykura C, Chao-Chu J, Schneider S, Dzhindzhev NS, Gottardo M, Riparbelli MG, et al. Conserved molecular interactions in centriole-to-centrosome conversion. Nat Cell Biol. 2016;18:87-99.
3. Tsuchiya Y, Yoshiba S, Gupta A, Watanabe K, Kitagawa D. Cep295 is a conserved scaffold protein required for generation of a bona fide mother centriole. Nat Commun. 2016;7:12567.
4. Dzhindzhev NS, Tzolovsky G, Lipinszki Z, Schneider S, Lattao R, Fu J, Debski J, Dadlez M, Glover DM. Plk4 phosphorylates Ana2 to trigger Sas6 recruitment and procentriole formation. Curr Biol. 2014;24:2526-2532.
5. Ohta M, Ashikawa T, Nozaki Y, Kozuka-Hata H, Goto H, Inagaki M, Oyama M, Kitagawa D. Direct interaction of Plk4 with STIL ensures formation of a single procentriole per parental centriole. Nat Commun. 2014;5:5267.
6. Tsou MF, Stearns T. Mechanism limiting centrosome duplication to once per cell cycle. Nature. 2006;442:947-951.
7. Wong C, Stearns T. Centrosome number is controlled by a centrosome-intrinsic block to reduplication. Nat Cell Biol. 2003;5:539-544.
8. Davoli T, de Lange T. The causes and consequences of polyploidy in normal development and cancer. Annu Rev Cell Dev Biol. 2011;27:585-610.
9. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang CZ, Wala J, Mermel CH, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet. 2013;45:1134-1140.
10. Krzywicka-Racka A, Sluder G. Repeated cleavage failure does not establish centrosome amplification in untransformed human cells. J Cell Biol. 2011;194:199-207.
11. Ganem NJ, Godinho SA, Pellman D. A mechanism linking extra centrosomes to chromosomal instability. Nature. 2009;460:278-282.
12. Galofre C, Asensio E, Ubach M, Torres IM, Quintanilla I, Castells A, Camps J. Centrosome reduction in newly-generated tetraploid cancer cells obtained by separase depletion. Sci Rep. 2020;10:9152.
13. Ganem NJ, Cornils H, Chiu SY, O'Rourke KP, Arnaud J, Yimlamai D, Thery M, Camargo FD, Pellman D. Cytokinesis failure triggers Hippo tumor suppressor pathway activation. Cell. 2014;158:833-848.
14. Fava LL, Schuler F, Sladky V, Haschka MD, Soratroi C, Eiterer L, Demetz E, Weiss G, Geley S, Nigg EA, et al. The PIDDosome activates p53 in response to supernumerary centrosomes. Genes Dev. 2017;31:34-45.
15. Burigotto M, Mattivi A, Migliorati D, Magnani G, Valentini C, Roccuzzo M, Offterdinger M, Pizzato M, Schmidt A, Villunger A, et al. Centriolar distal appendages activate the centrosome-PIDDosome-p53 signalling axis via ANKRD26. EMBO J. 2021;40:e104844.
16. Habedanck R, Stierhof YD, Wilkinson CJ, Nigg EA. The Polo kinase Plk4 functions in centriole duplication. Nat Cell Biol. 2005;7:1140-1146.
17. Kleylein-Sohn J, Westendorf J, Le Clech M, Habedanck R, Stierhof YD, Nigg EA. Plk4-induced centriole biogenesis in human cells. Dev Cell. 2007;13:190-202.
18. Marthiens V, Rujano MA, Pennetier C, Tessier S, Paul-Gilloteaux P, Basto R. Centrosome amplification causes microcephaly. Nat Cell Biol. 2013;15:731-740.
19. Kulukian A, Holland AJ, Vitre B, Naik S, Cleveland DW, Fuchs E. Epidermal development, growth control, and homeostasis in the face of centrosome amplification. Proc Natl Acad Sci U S A. 2015;112:E6311-E6320.
20. Vitre B, Holland AJ, Kulukian A, Shoshani O, Hirai M, Wang Y, Maldonado M, Cho T, Boubaker J, Swing DA, et al. Chronic centrosome amplification without tumorigenesis. Proc Natl Acad Sci U S A. 2015;112:E6321-E6330.
21. Sercin O, Larsimont JC, Karambelas AE, Marthiens V, Moers V, Boeckx B, Le Mercier M, Lambrechts D, Basto R, Blanpain C. Transient PLK4 overexpression accelerates tumorigenesis in p53-deficient epidermis. Nat Cell Biol. 2016;18:100-110.
22. Levine MS, Bakker B, Boeckx B, Moyett J, Lu J, Vitre B, Spierings DC, Lansdorp PM, Cleveland DW, Lambrechts D, et al. Centrosome amplification is sufficient to promote spontaneous tumorigenesis in mammals. Dev Cell. 2017;40:313-322.e5.
23. Schmidt TI, Kleylein-Sohn J, Westendorf J, Le Clech M, Lavoie SB, Stierhof YD, Nigg EA. Control of centriole length by CPAP and CP110. Curr Biol. 2009;19:1005-1011.
24. Chang J, Cizmecioglu O, Hoffmann I, Rhee K. PLK2 phosphorylation is critical for CPAP function in procentriole formation during the centrosome cycle. EMBO J. 2010;29:2395-2406.
25. Kong D, Sahabandu N, Sullenberger C, Vasquez-Limeta A, Luvsanjav D, Lukasik K, Loncarek J. Prolonged mitosis results in structurally aberrant and over-elongated centrioles. J Cell Biol. 2020;219:e201910019.
26. Kohlmaier G, Loncarek J, Meng X, McEwen BF, Mogensen MM, Spektor A, Dynlacht BD, Khodjakov A, Goncy P.

Overly long centrioles and defective cell division upon excess of the SAS-4-related protein CPAP. Curr Biol. 2009;19:1012-1018.
27. Marteil G, Guerrero A, Vieira AF, de Almeida BP, Machado P, Mendonca S, Mesquita M, Villarreal B, Fonseca I, Francia ME, et al. Over-elongation of centrioles in cancer promotes centriole amplification and chromosome missegregation. Nat Commun. 2018;9:1258.
28. Cabral G, Sans SS, Cowan CR, Dammermann A. Multiple mechanisms contribute to centriole separation in C. elegans. Curr Biol. 2013;23:1380-1387.
29. Seo MY, Jang W, Rhee K. Integrity of the pericentriolar material is essential for maintaining centriole association during M phase. PLoS One. 2015;10:e0138905.
30. Lee K, Rhee K. Separase-dependent cleavage of pericentrin B is necessary and sufficient for centriole disengagement during mitosis. Cell Cycle. 2012;11:2476-2485.
31. Kim J, Lee K, Rhee K. PLK1 regulation of PCNT cleavage ensures fidelity of centriole separation during mitotic exit. Nat Commun. 2015;6:10076.
32. Kim J, Kim J, Rhee K. PCNT is critical for the association and conversion of centrioles to centrosomes during mitosis. J Cell Sci. 2019;132:jcs225789.
33. Jung GI, Rhee K. Triple deletion of TP53, PCNT, and CEP215 promotes centriole amplification in the M phase. Cell Cycle. 2021;20:1500-1517.
34. Sala R, Farrell KC, Stearns T. Growth disadvantage associated with centrosome amplification drives population-level centriole number homeostasis. Mol Biol Cell. 2020;31:2646-2656.
35. Chiba S, Okuda M, Mussman JG, Fukasawa K. Genomic convergence and suppression of centrosome hyperamplification in primary p53-/- cells in prolonged culture. Exp Cell Res. 2000;258:310-321.
36. Quintyne NJ, Reing JE, Hoffelder DR, Gollin SM, Saunders WS. Spindle multipolarity is prevented by centrosomal clustering. Science. 2005;307:127-129.
37. Kwon M, Godinho SA, Chandhok NS, Ganem NJ, Azioune A, Thery M, Pellman D. Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. Genes Dev. 2008;22:2189-2203.
38. Leber B, Maier B, Fuchs F, Chi J, Riffel P, Anderhub S, Wagner L, Ho AD, Salisbury JL, Boutros M, et al. Proteins required for centrosome clustering in cancer cells. Sci Transl Med. 2010;2:33ra38.
39. Kwon M, Bagonis M, Danuser G, Pellman D. Direct microtubule-binding by Myosin-10 orients centrosomes toward retraction fibers and subcortical actin clouds. Dev Cell. 2015;34:323-337.
40. Baudoin NC, Nicholson JM, Soto K, Martin O, Chen J, Cimini D. Asymmetric clustering of centrosomes defines the early evolution of tetraploid cells. eLife. 2020;9:e54565.
41. Godinho SA, Picone R, Burute M, Dagher R, Su Y, Leung CT, Polyak K, Brugge JS, Thery M, Pellman D. Oncogene-like induction of cellular invasion from centrosome amplification. Nature. 2014;510:167-171.
42. Ganier O, Schnerch D, Oertle P, Lim RY, Plodinec M, Nigg EA. Structural centrosome aberrations promote non-cell-autonomous invasiveness. EMBO J. 2018;37:e98576.
43. Watanabe Y, Honda S, Konishi A, Arakawa S, Murohashi M, Yamaguchi H, Torii S, Tanabe M, Tanaka S, Warabi E, et al. Autophagy controls centrosome number by degrading Cep63. Nat Commun. 2016;7:13508.
44. Wu Q, Yu X, Liu L, Sun S, Sun S. Centrosome-phagy: implications for human diseases. Cell Biosci. 2021;11:49.

Beyond Conventional Fluorescence: Bringing Quantitative Microscopy into Routine Biological Research

Kyoohyun Kim

Core Facility Microscopy, Max-Planck-Zentrum für Physik und Medizin, 91054 Erlangen, Germany
kyoohyun.kim@mpzpm.mpg.de

Abstract. Recent advances in quantitative microscopy are transforming biological imaging from qualitative visualization toward quantitative characterization of cellular structure and function. While fluorescence microscopy remains indispensable for revealing molecular localization and dynamics, it provides limited information about the intrinsic physical properties of living cells. Label-free techniques such as optical diffraction tomography (ODT) and Brillouin microscopy address this limitation by providing complementary biophysical information. ODT reconstructs three-dimensional refractive index distributions to quantify cellular morphology, dry mass, and mass density, whereas Brillouin microscopy probes intracellular mechanical properties through spontaneous Brillouin scattering. Together, these complementary modalities provide an integrated framework for investigating the molecular, structural, and mechanical organization of biological systems without exogenous labels. As these technologies mature, improving their accessibility and integration into routine biological research will be essential for realizing the full potential of quantitative microscopy. This review highlights recent advances in ODT and Brillouin microscopy and discusses their growing roles in shaping the future of quantitative microscopy.

Keywords: Optical diffraction tomography, Brillouin microscopy, Core facility

1. Introduction

Since the invention of the optical microscope more than four centuries ago, microscopy has continuously transformed our understanding of biological systems. Bright-field microscopy first revealed cellular morphology, while fluorescence microscopy revolutionized modern biology by enabling visualization of specific proteins and intracellular structures. Today, fluorescence microscopy has become an indispensable tool across the life sciences, providing unparalleled molecular specificity for studying cellular organization and signaling.

Despite these advances, many biological questions remain difficult to answer using fluorescence imaging alone. While fluorescence microscopy reveals the spatial distribution of labeled molecules, it provides only indirect information about the intrinsic physical state of cells. Cellular physiology is also governed by biophysical properties such as intracellular density, macromolecular crowding, and mechanical stiffness, which change during development, differentiation, immune activation, aging, and disease progression. These properties are increasingly recognized as important functional biomarkers.

The growing interest in mechanobiology and quantitative cell biology has therefore driven the development of label-free imaging techniques capable of measuring these intrinsic physical properties. These approaches provide complementary information that cannot be obtained from molecular imaging alone, reflecting a broader evolution of microscopy from visualizing cellular structures to quantitatively characterizing biological systems.

Among these emerging techniques, optical diffraction tomography (ODT) and Brillouin microscopy have attracted particular attention. Although based on distinct physical principles, both exploit light–matter interactions to provide quantitative information inaccessible to conventional microscopy. ODT reconstructs the three-dimensional refractive index (RI) distribution of cells and tissues, enabling quantitative analysis of cellular morphology and mass density (protein concentration), whereas Brillouin microscopy maps intracellular mechanical properties through spontaneous Brillouin scattering.This review highlights recent advances in ODT and Brillouin microscopy and discusses how these technologies are shaping the future of quantitative microscopy and facilitating their broader adoption in biological research.

2. ODT for quantifying mass density

ODT is a three-dimensional quantitative phase imaging technique conceptually analogous to X-ray computed tomography (CT). By acquiring multiple quantitative phase images under different illumination angles, ODT reconstructs the intrinsic three-dimensional RI distribution of cells and tissues (Fig. 1A). Because the RI of most biological materials is proportional to the local concentration of intracellular biomolecules, reconstructed RI maps provide not only high-resolution morphological images (~100 nm) but also quantitative measurements of volume, dry mass, and mass density[1].

A major strength of ODT is its versatility. It has been successfully applied to a wide range of biological specimens, including red blood cells, immune cells, neurons, microorganisms, organoids, and early embryos (Fig. 1B-E). By visualizing subtle variations in RI and mass density at subcellular resolution, ODT provides quantitative biophysical information and enables the development of physical models for understanding cellular organization and function[2–4].Among intracellular organelles, the nucleus provides an illustrative example of the biological insights enabled by ODT. Despite its high concentration of genetic material, ODT has consistently shown that the nucleoplasm possesses a lower mass density than the surrounding cytoplasm across multiple cell types and developmental stages (Fig. 1B, D, and E). This conserved phenomenon highlights the ability of quantitative microscopy to uncover previously unrecognized physical principles governing cellular organization.

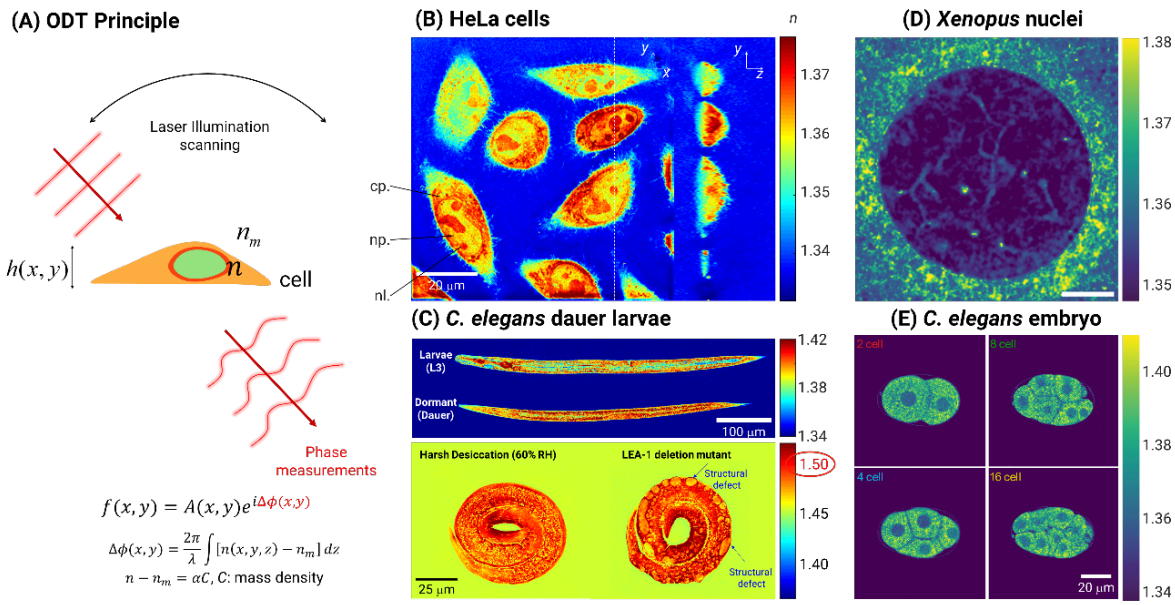


Fig. 1. (A) Principle of ODT. (B) Cross-sectional RI tomogram of HeLa cells showing cytoplasm (cp.), nucleoplasm (np.), and nucleoli (adapted from [2]). (C) Cross-sectional RI tomograms of *C. elegans* larvae at the L3 larval stage, dauer stage, and desiccated dauer in control and *lea-1* deletion mutants (adapted from [3]). (D) Cross-sectional RI tomogram of *Xenopus* nuclei and (E) *C. elegans* embryo demonstrating the lower mass density of nucleus than cytoplasm (adapted from [4])

3. Brillouin microscopy for quantifying mechanical properties

Cellular and tissue mechanics play a fundamental role in numerous biological processes, including embryonic development, tissue morphogenesis, stem cell differentiation, immune-cell activation, cancer invasion, and neurodegeneration. Understanding these processes therefore requires imaging techniques capable of probing mechanical properties with high spatial resolution while preserving the native physiological state of living specimens.

Brillouin microscopy has emerged as one of the most promising label-free approaches for biomechanical imaging. Brillouin microscopy measures the frequency shift of light inelastically scattered by thermally excited acoustic phonons within biological materials (Fig. 2A). The measured Brillouin frequency shift is related to the longitudinal modulus (a proxy of inverse of the compressibility) of the material, providing an

intrinsic optical contrast associated with its mechanical properties[5].Unlike conventional mechanical characterization techniques such as atomic force microscopy (AFM) or micropipette aspiration, Brillouin microscopy is entirely non-contact and can probe mechanical properties deep within living tissues. This capability has opened new opportunities for investigating biomechanics in intact biological specimens, such as cultured cells, regenerating spinal cord, corneal tissue, cartilage, and brain tissue (Fig. 2B-D).

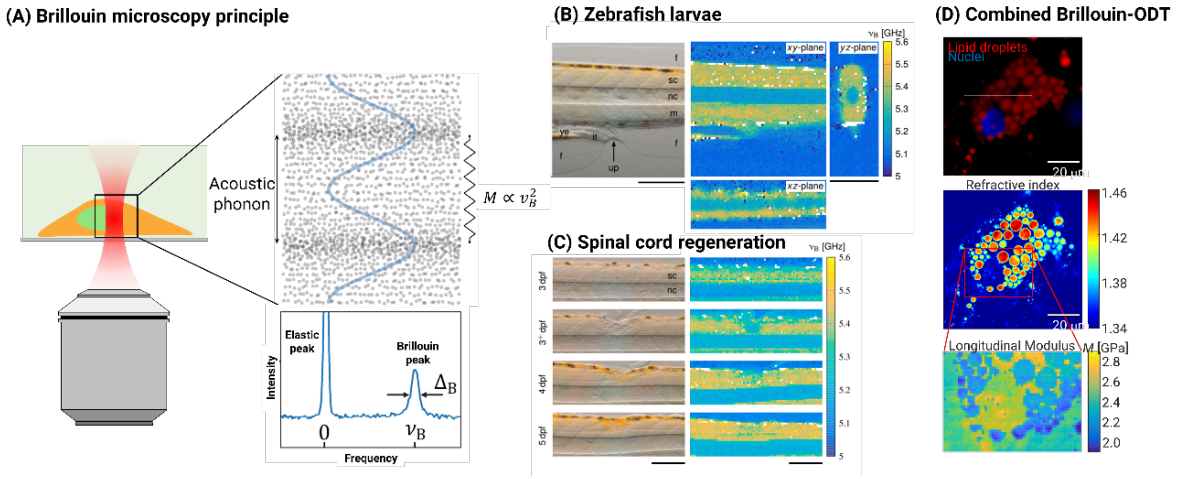


Fig. 2. (A) Principle of Brillouin microscopy (adapted from [6]). (B) Brightfield image and corresponding Brillouin frequency map of a zebrafish larva. (C) Progression of spinal cord regeneration in a zebrafish larva following injury (adapted from [5]). (D) Combined fluorescence microscopy (top), ODT (middle), and Brillouin microscopy (bottom) of an adipocyte, together with the corresponding longitudinal modulus map(adapted from [7]).

4. Towards Integrated Quantitative Microscopy in Core Facility

Neither optical nor mechanical measurements alone can fully describe the complexity of living systems. Fluorescence microscopy provides molecular specificity, ODT quantifies cellular architecture and mass density, and Brillouin microscopy measures mechanical properties. Together, these complementary techniques enable a more comprehensive understanding of biological systems than any individual modality alone. Furthermore, because the Brillouin frequency shift depends on the RI and mass density of the sample, ODT provides complementary information for a more quantitative interpretation of Brillouin measurements[7] (Fig. 2D).

Despite their considerable potential, ODT and Brillouin microscopy remain less accessible than conventional fluorescence microscopy. As these technologies move from specialized research laboratories to shared imaging facilities, successful implementation extends beyond instrument development. For successful implementation, researchers need to identify biological questions that uniquely benefit from quantitative optical and mechanical measurements, requiring close collaboration between imaging specialists and life scientists. At the same time, standardized operating procedures (SOPs), structured user training, and intuitive analysis workflows are essential to enable researchers without expertise in optical engineering to confidently adopt these technologies.

Our experience at the Core Facility Microscopy of the Max-Planck-Zentrum für Physik und Medizin (MPZPM) illustrates how accessibility can be improved through dedicated user support. Through comprehensive user training, SOPs, and close scientific consultation, researchers from diverse disciplines have been able to independently operate advanced quantitative microscopy systems and incorporate these technologies into their research. This experience highlights the increasingly important role of microscopy core facilities in broadening access to quantitative microscopy and supporting interdisciplinary biological research.

References

1. Kim, G. et al.: Holotomography. Nat. Rev. Methods Prim. 4(1), 51 (2024).
2. Kim, K., Guck, J.: The Relative Densities of Cytoplasm and Nuclear Compartments Are Robust against Strong Perturbation. Biophys. J. 119(10), 1946–1957 (2020).
3. Kim, K., Gade, V. R., Kurzchalia, T. V., Guck, J.: Quantitative imaging of Caenorhabditis elegans dauer larvae during cryptobiotic transition. Biophys. J. 121(7), 1219–1229 (2022).
4. Biswas, A., Muñoz, O., Kim, K. et al.: Conserved nucleocytoplasmic density homeostasis drives cellular organization across eukaryotes. Nat. Commun. 16, 7597 (2025).
5. Schlüßler, R., Möllmert, S. et al.: Mechanical Mapping of Spinal Cord Growth and Repair in Living Zebrafish Larvae by Brillouin Imaging. Biophys. J. 115(5), 911–923 (2018).
6. Beck, T. et al.: Optical characterization of molecular interaction strength in protein condensates. Mol. Biol. Cell 35, ar154 (2024).
7. Schlüßler, R., Kim, K. et al.: Correlative all-optical quantification of mass density and mechanics of subcellular compartments with fluorescence specificity. eLife 11, e68490 (2022).

From Cell Fate Curiosity to Everyday Genomics

Rinho Kim

Max Planck Institute of Biochemistry, Am Klopfspitz 18, 82152 Martinsried near Munich, Germany
rkim@biochem.mpg.de

Abstract. My path into research began with scientific curiosity about cell fate specification, first as an undergraduate in South Korea, driven by a desire to pursue independent research. That desire was realized during doctoral training on intestinal stem cell fate specification in the United States, yet even then I wondered whether a principal investigator career was the right fit for me. This uncertainty led me to try different paths in graduate school, and through the rejections I faced, I learned what I could do well. Those experiences became the foundation for where I am today: leading an NGS core facility in Germany for the past seven years, in the middle of a field undergoing rapid technical advancement and an ever-widening range of real-world applications. If there is one thing I would pass on to younger scientists searching for their own path, it is this: try things out without fear, and stay open to the people around you who are quietly pointing the way.

Keywords: Career development, sequencing technology, genomic data security

1. From Curiosity to Technical Expertise

My first taste of research came as an undergraduate research assistant in South Korea, working on a project that used proteomic approaches to distinguish satellite cells from the myoblasts they were becoming. What stayed with me was not the technique itself but the underlying puzzle: cells carrying identical genetic information were nonetheless committing to entirely different fates. That curiosity followed me into my undergraduate years, and when a senior advised me that if I was going to end up in the United States eventually anyway, I should go while still young rather than wait, I took that advice, and, with only a bachelor's degree in hand, applied for a research assistant position at Brandeis University and joined the lab's research there. Under a supportive supervisor, I became a co-author on my first paper, on how a single homeodomain transcription factor determines sensory neuron identity in C. elegans. Seeing, in real data, how one transcription factor could decide a cell's identity is what made me fall for genetics research, and it opened the door to a PhD at the University of Pennsylvania's medical school.

There, working with a mouse model, I studied the same underlying question, this time through the epigenetic regulation of intestinal stem cells, examining how Tet1-mediated DNA hydroxymethylation shaped the balance between stem cells and their differentiated descendants. On the advice of several seniors, who saw next-generation sequencing (NGS) as a technology about to expand rapidly, I joined a lab built around it. Yet even while doing the independent research I had long wanted, I found I could not quite picture myself as a principal investigator, watching how many talented peers were converging on that same narrow path. Uncertain what fit me better, I tried strategy consulting, and was turned down everywhere I applied. Pointed back toward life sciences industry roles, I was instead advised by another senior to try a short postdoc first, and chose to do it somewhere unexpected: Munich, Germany.

My postdoctoral project at Helmholtz Munich applied epigenetic analysis to finding biomarkers in diabetes and obesity patient samples, working with very limited sample amounts. What stood out was not the technical difficulty but a clearer realization: I was more energized advising other labs on their NGS work than leading my own funded research. Barely six months in, I concluded the principal investigator path was not for me. I began building industry-relevant skills, and it was during that search that I found an advertisement for an NGS specialist and core facility head at the Max Planck Institute of Biochemistry, a role that seemed designed exactly for me.

2. Seven Years of Facility Management, and a Changing Sequencing Landscape

I was fortunate from the very start: the team I inherited was already skilled and well trained, giving me a stable foundation. With the institute's full support, I expanded the facility over the following years, acquiring high-end instruments, growing the group with strong PhD-level scientists, and, more recently, adding a dedicated bioinformatics specialist. That small team has since grown into a group made up entirely of tenured, highly capable people, one now known across the wider Max Planck community as a benchmark NGS core.

NGS technology has advanced remarkably fast over the past two decades or so, and it is now widely used in everyday life. Clinical cancer diagnostics, non-invasive prenatal and newborn screening, and direct-to-consumer ancestry testing have each followed a similar arc: introduced as niche or high-risk applications between 2005 and 2012, then expanding rapidly to become routine or mainstream within roughly a decade [1-3]. NGS-based testing in non-small-cell lung cancer now exceeds 90 percent adoption in many settings, universal prenatal screening is standard practice in numerous countries, and tens of millions of people worldwide have taken a consumer ancestry test. Figure 1 traces this two-decade shift; the three curves are each normalized to their own current scale, so they compare the timing and steepness of adoption rather than the absolute size of each application.

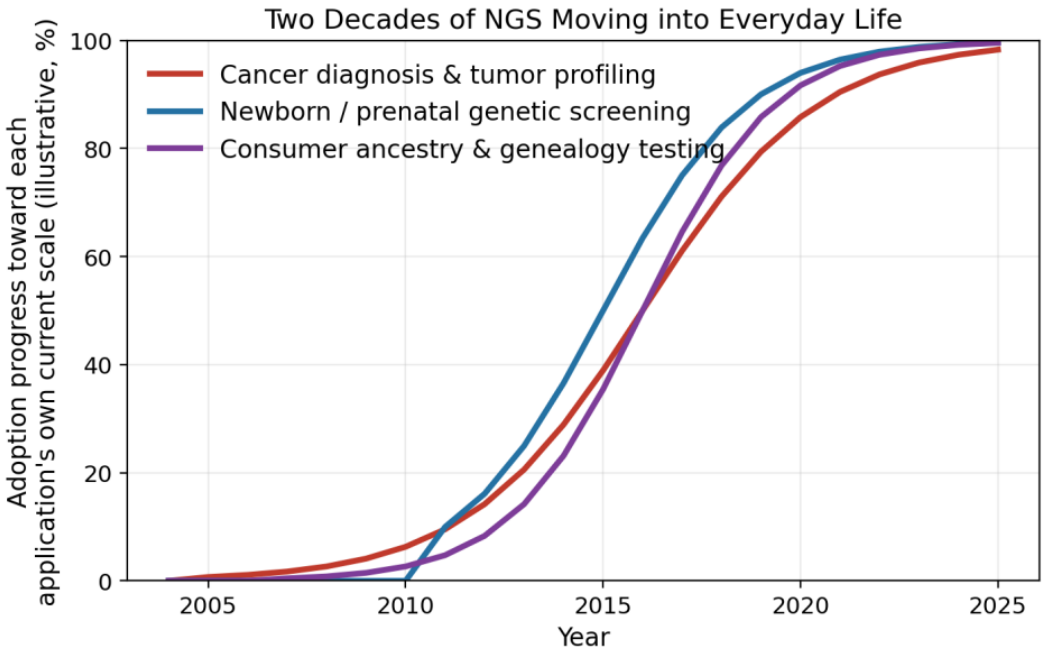


Fig. 1. Illustrative adoption curves for three NGS-enabled applications in everyday medicine and consumer life, 2004-2025, each independently normalized to its own current scale and anchored to known introduction and adoption milestones [1][2][3].

Running a core facility through this period has also meant watching one of the most consequential shifts in the sequencing supply market in over a decade. For most of the 2010s and early 2020s, Illumina's short-read platforms held a near-monopoly position, and the NovaSeq X remains by far the most widely adopted platform across core facilities worldwide [4]. This dominance rested partly on foundational patents covering reversible-terminator sequencing chemistry, and as those patents have expired, competitors have moved quickly to fill the space. Element Biosciences launched its benchtop AVITI24 system in 2024, whose key strength lies in generating multiomic data including genomics and targeted RNA/protein detections, all from a single run [5]. In June 2026, Roche entered the field with its Axelios 1 platform, which uses a fundamentally different sequencing-by-expansion chemistry and can complete a run in as little as 15 minutes, fast enough to support same-day clinical diagnostics [6]. MGI's DNBSEQ-T7+, meanwhile, has carved out its own niche in

large-scale projects, offering the highest sustained daily throughput in the market at a notably competitive price [7]. That same month, Samsung Electronics deepened its position in this market. It invested an additional 175 million US dollars in Element Biosciences and became the company's largest shareholder. The move signals Samsung's intent to combine Element's sequencing technology with its own artificial intelligence and diagnostics capabilities [8][9].

Table 1. Snapshot of leading short-read/next-generation sequencing platforms as of mid-2026.

Platform	Max Output per Run	Typical Run Time	Distinguishing Throughput Trait
Illumina NovaSeq X Plus	Up to 21 Tb	13–48 hrs	Highest raw output ceiling[4]
Element AVITI24	Up to 1.5 Tb	24–38 hrs	Strong single-cell/multiomic capacity [5]
Roche Axelios 1	1.8 Tb	15 min–4 hrs	Fastest turnaround per genome [6]
MGI DNBSEQ-T7+	Up to 14 Tb	20–30 hrs	Highest sustained daily throughput; now facing US market restrictions [7]

3. What the Next Chapter Demand

We are now in an era where artificial intelligence is being deployed ever more broadly and evolving faster than most could have predicted, which alone makes the security of genomic data an increasingly urgent concern. The same period has also made large-scale DNA synthesis and CRISPR-based editing routinely accessible, raising the stakes even further: genomic data is no longer just something to be read and analyzed, but something that can, in principle, be rewritten. This growing emphasis on speed and throughput has not stayed separate from geopolitics. In May 2026, Illumina's CEO Jacob Thaysen joined President Trump's business delegation to the US-China summit in Beijing, a visit notable precisely because sequencing had, by then, become a strategic technology rather than a purely commercial one [10]. Around the same period, the United States finalized rules under Executive Order 14117 restricting the transfer of bulk human genomic data to China and other countries of concern, while the BIOSECURE Act moved to bar federal contracts involving Chinese genomics firms, including BGI, MGI, and Complete Genomics, on national security grounds [11][12]. For a core facility, this is a reminder that platform choice is no longer a purely technical decision, and that securing this data, and under what rules, deserves at least as much attention as run time or cost per genome.

What I find most rewarding about a life spent in scientific research is precisely that it never lets me settle into a stagnant or repetitive routine. So, if there is one thing I would pass on to younger scientists still searching for their own path, it is this: try things out without fear, and stay open to the people around you who are quietly pointing the way.

References

1. Karlovich, C.A., Williams, P.M.: Clinical Applications of Next-Generation Sequencing in Oncology. PMC (2019).
2. Jayashankar, S.S., et al.: Non-Invasive Prenatal Testing (NIPT): Reliability, Challenges. PMC (2023).
3. Global Market Insights: Direct-to-Consumer (DTC) Genetic Testing Market Report (2025).
4. Illumina: NovaSeq X Series specifications. Illumina, Inc. (2023).

5. Element Biosciences: AVITI24 System and Specifications. (2024).
6. Roche Diagnostics: AXELIOS 1 Platform for Next-Generation Sequencing, spec sheet. (2026).
7. MGI Tech / GenomeWeb: MGI Tech Launches DNBSEQ-T7+ Sequencer. (2025).
8. Reuters: Samsung Elec invests $175 million to become top shareholder in Element Biosciences. (2026-06-09).
9. Samsung Newsroom Korea: 삼성전자, 美 '엘리먼트바이오사이언스'社 1대주주지위확보. (2026-06-10).
10. Thaysen, J.: LinkedIn post on joining the U.S. business delegation to the U.S.-China Summit in Beijing. (2026-05-19).
11. U.S. Department of Justice / Federal Register: Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern, Final Rule. (2025-01-08).
12. Science.org: Congress close to passing law that would freeze out certain Chinese biotechs (BIOSECURE Act). (2025-10-27). arlovich, C.A., Williams, P.M.: Clinical Applications of Next-Generation Sequencing in Oncology. PMC (2019).

Technologies for Reproductive Capacity Regulation in Humans and Animals

Youn-Su Lee[1]

[1]Department of Pet Health, Kwangju Women's University, 40 Gwangjuyeodae-gil, Gwangsan-gu, Gwangju 62396, Republic of Korea
yslee@kwu.ac.kr

Abstract. Reproductive-capacity technologies are expanding rapidly, but human and animal applications should not be presented as a single continuous intervention pathway. In humans, the major aims are clinical fertility care, fertility preservation, prevention of heritable disease, reversible contraception, reproductive autonomy, and long-term safety for offspring. In animals, the major aims are livestock breeding, genetic dissemination, companion-animal fertility control, animal welfare, biodiversity protection, and ecological containment. This review separates these domains while comparing their shared technical logic: gametes, embryos, endocrine signaling, germline transmission, and population output are becoming measurable and adjustable. The central trend is therefore not simply greater control, but more domain-specific control with different evidence standards, ethical obligations, and governance thresholds.

Keywords: Reproductive capacity regulation; Fertility preservation; Male contraception; Animal reproductive biotechnology

1. Human reproductive technologies

Human reproductive technologies are clinical tools that act at several separate points: gamete production, embryo creation, embryo selection, uterine transfer, fertility preservation, and contraception. The purpose is not to maximize reproduction in the abstract, but to support patient-specific goals under medical, ethical, and regulatory constraints. This distinction matters because a technology that increases the chance of pregnancy in one group may add cost, intervention burden, or uncertainty in another. Evidence standards therefore have to be tied to indications such as infertility diagnosis, age, cancer treatment, mitochondrial disease risk, or contraceptive need.

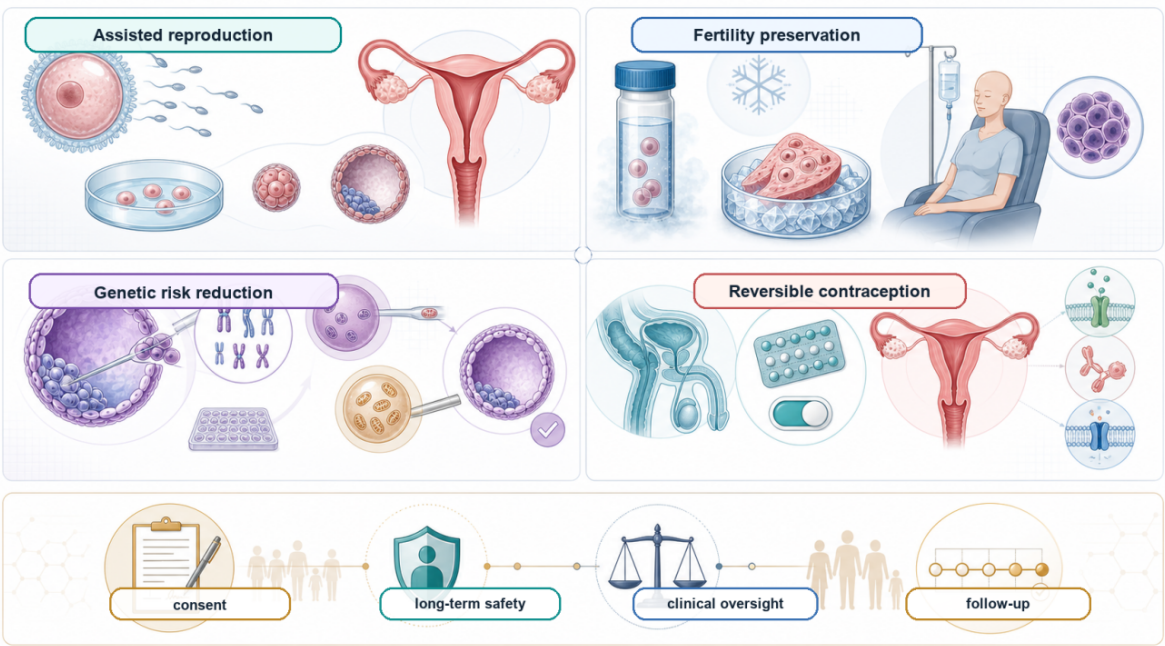


Fig. 1. Human reproductive-capacity technologies.

1.1 Assisted reproduction and embryo assessment

Assisted reproductive technology has shifted from a procedure-centered model to a risk-stratified model. Conventional IVF and intracytoplasmic sperm injection create embryos; embryo culture, blastocyst assessment, cryopreservation, and single-embryo transfer then determine how many embryos are transferred and when. Preimplantation genetic testing for aneuploidy (PGT-A) is a clear example of the need for careful interpretation. The large randomized trial by Yan and colleagues showed that PGT-A cannot be assumed to improve live birth in every patient group [1]. Its value depends on maternal age, embryo number, laboratory performance, mosaicism interpretation, and whether the outcome is per-transfer success, cumulative live birth, miscarriage reduction, or time to pregnancy.

A second human technology axis is prevention of severe inherited disease. Mitochondrial donation is not an embryo selection tool in the ordinary sense; it changes the cytoplasmic genetic context so that people carrying pathogenic mitochondrial DNA variants can attempt genetically related reproduction with reduced transmission risk. The 2025 clinical report combining mitochondrial donation with preimplantation genetic testing for mtDNA disease is important because it moves the field from proof-of-principle toward monitored clinical outcomes [2]. At the same time, the technology raises questions about residual mutant mtDNA carryover, nuclear-mitochondrial compatibility, embryo loss, and the need for child and possibly intergenerational follow-up.

1.2 Fertility preservation and germline modeling

Fertility preservation has become a standard part of cancer survivorship and other gonadotoxic-treatment planning. Embryo cryopreservation is mature but requires sperm and time for ovarian stimulation. Oocyte cryopreservation gives more reproductive autonomy but remains age- and ovarian-reserve-dependent. Ovarian tissue cryopreservation can be used when treatment is urgent or the patient is prepubertal, but transplantation introduces questions about graft longevity, ischemic follicle loss, and possible malignant-cell reintroduction in some cancers. The systematic review and meta-analysis by Fraison and colleagues summarize live-birth outcomes across embryo, oocyte, and ovarian tissue cryopreservation, making clear that preservation is not one technique but a menu of context-dependent options [3].

Human in vitro gametogenesis and germline modeling sit at a different translational stage. They are often discussed as future fertility-restoration technologies, but their immediate value is mechanistic. Reconstitution of human germline epigenetic reprogramming in vitro provides a way to study how primordial germ-cell-like cells erase and rebuild epigenetic marks, how meiotic competence might emerge, and where infertility-causing errors occur [4]. For clinical translation, the unresolved issues are substantial: imprinting fidelity, chromosomal segregation, epigenetic memory, culture-induced mutation, and whether in vitro-derived gametes can ever meet a safety threshold appropriate for human reproduction.

1.3 Reversible contraception and male-targeted innovation

Contraception regulates reproductive capacity by preventing conception rather than enabling it, and the human design goal is usually reversibility. Recent research is especially active in male contraception because the method portfolio has remained narrow for decades. Acute inhibition of soluble adenylyl cyclase is attractive because it acts downstream of sperm production, transiently blocking sperm motility and capacitation in preclinical models [5]. This suggests an on-demand strategy that would differ from hormonal suppression of spermatogenesis. However, clinical translation has to prove rapid onset, reliable duration, full reversibility, lack of systemic toxicity, and a failure rate acceptable to users and partners.

YCT-529 represents a complementary approach: an oral nonhormonal antagonist of retinoic acid receptor alpha that is intended to suppress male fertility by interfering with a pathway required for spermatogenesis [6]. Compared with hormonal male contraception, nonhormonal approaches may avoid testosterone-suppression

side effects, but they still face difficult development criteria. A clinically useful male contraceptive must preserve libido and general health, avoid long-term germ-cell damage, recover predictably after discontinuation, and be acceptable in real-world use. Thus, human fertility regulation is increasingly molecular, but the bar for adoption remains behavioral and social as much as pharmacological.

1.4 Clinical implementation priorities

Across human technologies, implementation depends on outcome definitions. A clinic, patient, and regulator may each value a different endpoint: fertilization rate, embryo yield, euploid embryo number, miscarriage reduction, cumulative live birth, time to pregnancy, avoidance of disease transmission, or contraceptive continuation. Short-term laboratory success can therefore be misleading if it is not tied to patient-centered outcomes. The field is moving toward more complete reporting of cumulative outcomes, adverse events, embryo attrition, and follow-up beyond birth.

Equity is also part of technical evaluation. Fertility preservation, PGT, mitochondrial donation, and emerging male contraception can widen reproductive options, but only if access, counseling quality, cost, and regulatory availability are addressed. A technology that is biologically elegant but available only to a narrow patient group may deepen rather than reduce reproductive inequity. For this reason, human reproductive-capacity control should be evaluated as a combined clinical, ethical, and health-system problem.

1.5 Data infrastructure and follow-up

A central implementation gap is longitudinal data. Assisted reproduction, fertility preservation, mitochondrial donation, and male contraception all require follow-up intervals that are longer than the laboratory or early clinical trial window. For ART and PGT, the relevant denominator is often not one embryo transfer but the whole treatment episode: stimulation, embryo attrition, cryostorage, transfer, miscarriage, live birth, neonatal outcome, and later child health. For fertility preservation, the denominator is broader still, because many patients store gametes or tissue without returning to use them. Without registries that connect treatment exposure to later outcomes, success can be overestimated and rare safety signals can be missed.

The same point applies to emerging contraception. A male contraceptive may look promising in pharmacodynamic assays, but real-world value depends on adherence, partner trust, recovery after discontinuation, user-reported side effects, and pregnancy rates over typical rather than perfect use. Human reproductive technologies therefore need a mixed evidence model: mechanistic assays for biological plausibility, randomized or prospective clinical studies for efficacy, registries for safety, and qualitative work for acceptability. This is why translation should be treated as a clinical ecosystem rather than as a single device, assay, or drug moving linearly from bench to bedside.

2. Animal reproductive technologies

Animal reproductive technologies form a separate domain because the target unit is often not an individual patient but a herd, breed, colony, population, or ecosystem. The goals include accelerating genetic gain, disseminating valuable germlines, controlling sex ratio, reducing unwanted reproduction, protecting endangered populations, or suppressing invasive species. The ethical structure is therefore different from human medicine. Consent is not the organizing principle; animal welfare, stewardship, ownership, biodiversity, containment, and public legitimacy are central.

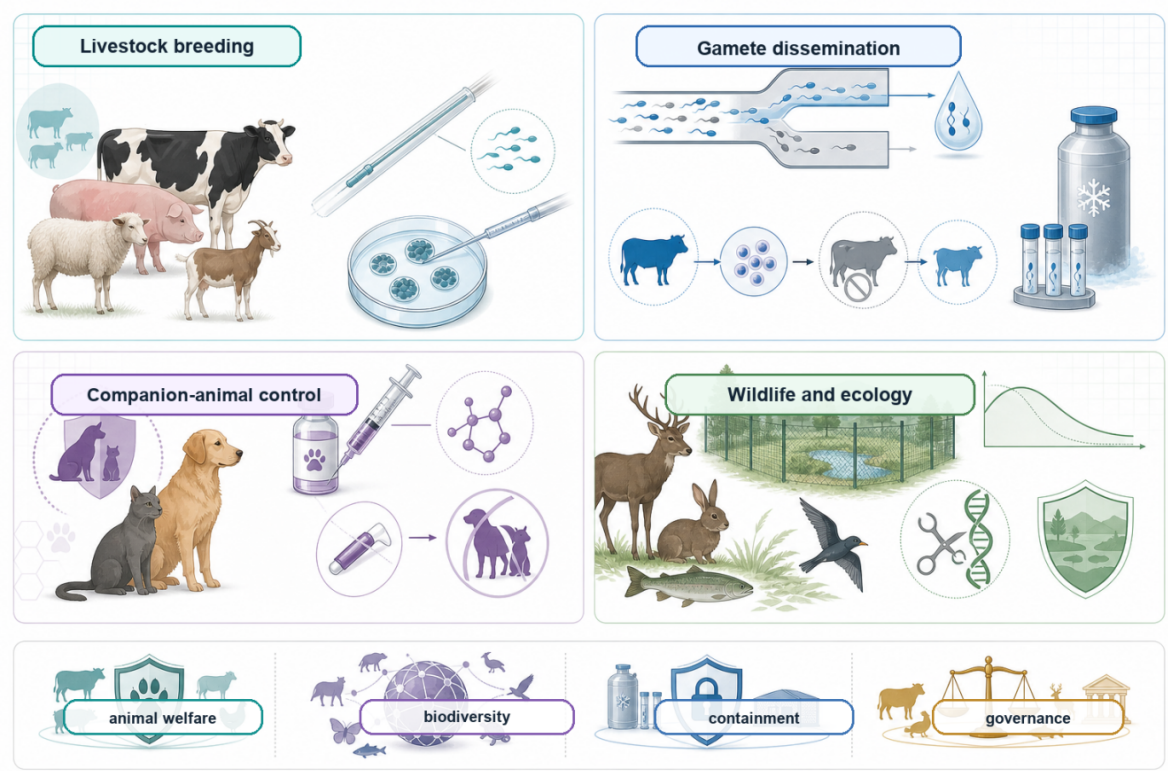


Fig. 2. Animal reproductive-capacity technologies.

2.1 Livestock breeding and gamete dissemination

In livestock, reproductive technologies regulate capacity by changing which gametes are used and how efficiently they produce offspring. Artificial insemination, embryo transfer, semen cryopreservation, genomic selection, and sexed semen already allow breeding programs to separate genetic merit from natural mating. Sex-sorted semen is a practical example: it can bias offspring sex toward dairy heifers or production goals, but the meta-analysis by Reese and colleagues shows that reproductive success after sorting can be reduced depending on sperm dose, sorting damage, insemination protocol, and female fertility status [10]. This is a reminder that precision at the gamete level does not automatically translate into equal fertility at the herd level.

Surrogate-sire technology pushes the separation of genotype and reproductive body even further. Donor-derived spermatogenesis after stem-cell transplantation into sterile NANOS2 knockout males demonstrates that a recipient male can function as a biological vehicle for another germline [9]. In principle, this could disseminate elite genetics, support breeding in species where artificial insemination is inefficient, or help conserve endangered germlines. In practice, it raises questions about gene editing of recipient animals, sterility reliability, germ-cell competition, biosecurity, welfare of surrogate males, and control over the distribution of valuable genetic lines.

2.2 Companion animals and non-surgical fertility control

Companion-animal fertility control has a different logic from livestock breeding. The objective is often to reduce unwanted births, shelter burden, and free-roaming population growth while minimizing repeated capture and surgery. Viral-vectored delivery of a feline anti-Mullerian hormone transgene produced durable contraception in female domestic cats [8]. This is scientifically important because it suggests a long-acting biologic route to fertility suppression. Yet the translational questions are not only whether pregnancy is prevented. A field-ready method must be safe across age and reproductive status, deliverable at scale, affordable, trackable, and acceptable to veterinarians, owners, shelters, and communities.

The companion-animal setting also highlights the difference between individual veterinary treatment and population management. A single animal may benefit from avoiding surgery, but population-level effects depend on coverage, duration, immigration, abandonment, and survival. Long-acting contraception can reduce births only if enough animals are reached and if fertility suppression lasts long enough to alter population dynamics. Therefore, the evaluation of these technologies needs reproductive endpoints, welfare endpoints, and ecological or demographic models.

2.3 Wildlife, gene drives, and ecological containment

Wildlife fertility control is the highest-governance animal domain because interventions can affect ecosystems beyond the treated animals. Immunocontraception, habitat management, and targeted fertility suppression are already used in some contexts, but gene-editing and inheritance-biasing systems raise a different level of concern. Super-Mendelian inheritance mediated by CRISPR-Cas9 in the female mouse germline demonstrated that mammalian gene-drive-like behavior is technically possible [11]. That proof does not mean open release is near or desirable. It means that ecological containment, molecular reversibility, species specificity, resistance evolution, and governance must be developed in parallel with the molecular tools.

Animal models also remain essential for understanding gametogenesis itself. The generation of functional oocytes from male mice in vitro is not a livestock management tool by itself, but it shows how animal systems can uncover the plasticity and constraints of germ-cell development [7]. Such studies inform reproductive biology, infertility research, and safety assessment. They should be interpreted within their species context: a mouse result can reveal mechanisms and risks without implying direct translation to human reproduction or immediate use in farm animals.

2.4 Animal implementation priorities

Animal translation is shaped by delivery logistics as much as by molecular potency. In livestock, the intervention has to fit breeding seasons, semen handling, veterinary labor, herd records, disease-control rules, and producer economics. In companion animals, a long-acting contraceptive has to work across variable health status, ownership status, and capture opportunities. In wildlife, even a highly specific fertility intervention may fail if immigration, compensatory survival, or altered behavior offsets reduced reproduction. These settings require population-level endpoints in addition to individual reproductive endpoints.

Welfare monitoring is therefore not an accessory to animal reproductive technology; it is part of the evidence base. A method that suppresses fertility but increases pain, social disruption, endocrine disease, or ecological imbalance would not count as successful regulation. Similarly, livestock tools that accelerate genetic gain can narrow genetic diversity if breeding decisions are too concentrated, while wildlife tools can shift ecosystem relationships if target populations decline quickly. The practical research agenda is to pair fertility outcomes with welfare scoring, genetic diversity metrics, biosecurity review, and ecological surveillance.

3. Conclusion

The safest framing is parallel rather than continuous. Human reproductive technologies should be judged by clinical benefit, autonomy, equity, heritable-risk management, informed consent, and offspring follow-up. Animal reproductive technologies should be judged by welfare, genetic diversity, containment, ecological impact, ownership, and stakeholder governance. Both domains are becoming more precise because they share molecular tools, but their acceptable endpoints are different. A pregnancy, a calf, a sterilized cat, and a suppressed invasive population are not interchangeable outcomes, even when the underlying biology involves gametes or embryos.

Near-term research should therefore prioritize domain-specific evidence. Human studies need prospective registries, transparent reporting of cumulative live birth, miscarriage, congenital outcomes, and long-term child health. Animal studies need fertility endpoints, welfare monitoring, genetic-diversity metrics, and

ecological modeling. Across both domains, the most important technical trend is controllability: the ability to tune duration, target tissue, target species, inheritance, and reversibility. The most important social trend is accountability: the requirement to show who benefits, who bears risk, and who has authority over deployment.

For human applications, governance begins before deployment: informed consent, counseling about uncertainty, limits on embryo testing claims, and postnatal follow-up are integral parts of the technology. For animal applications, governance extends outward: a breeding intervention may affect market structure and genetic diversity, while a wildlife fertility intervention may affect predators, competitors, vegetation, or disease ecology. The same molecular intervention can therefore require very different oversight depending on whether it is used in a clinic, a farm, a shelter, or an open landscape.

A useful practical framework is to ask four questions before translation. First, is the intervention reversible at the scale where harm could occur? Second, is the target specific enough to avoid off-target reproductive or ecological effects? Third, are monitoring systems strong enough to detect failure after deployment? Fourth, do affected communities or stakeholders have a meaningful role in deciding whether the intervention is acceptable? These questions help keep human and animal pathways separate while still allowing scientific comparison across technologies.

References

1. Yan J, Qin Y, Zhao H, et al. Live Birth with or without Preimplantation Genetic Testing for Aneuploidy. N Engl J Med. 2021;385(22):2047-2058. PMID: 34818479. DOI: 10.1056/NEJMoa2103613.

2. Hyslop LA, Blakely EL, Aushev M, et al. Mitochondrial Donation and Preimplantation Genetic Testing for mtDNA Disease. N Engl J Med. 2025;393(5):438-449. PMID: 40673696. DOI: 10.1056/NEJMoa2415539.

3. Fraison E, Huberlant S, Labrune E, et al. Live birth rate after female fertility preservation for cancer or haematopoietic stem cell transplantation: a systematic review and meta-analysis of embryo, oocyte and ovarian tissue cryopreservation. Hum Reprod. 2023;38(3):489-502. PMID: 36421038. DOI: 10.1093/humrep/deac249.

4. Murase Y, Yokogawa R, Yabuta Y, et al. In vitro reconstitution of epigenetic reprogramming in the human germ line. Nature. 2024;631(8019):170-178. PMID: 38768632. DOI: 10.1038/s41586-024-07526-6.

5. Balbach M, Rossetti T, Ferreira J, et al. On-demand male contraception via acute inhibition of soluble adenylyl cyclase. Nat Commun. 2023;14(1):637. PMID: 36788210. DOI: 10.1038/s41467-023-36119-6.

6. Shi R, Al Noman MA, Mannowetz N, et al. From Discovery to Clinical Trial: YCT-529, an Oral NonHormonal Male Contraceptive Targeting the Retinoic Acid Receptor Alpha. J Med Chem. 2026;69(2):1568-1605. PMID: 41524264. DOI: 10.1021/acs.jmedchem.5c03051.

7. Murakami K, Hamazaki N, Hamada N, et al. Generation of functional oocytes from male mice in vitro. Nature. 2023;615(7954):900-906. PMID: 36922585. DOI: 10.1038/s41586-023-05834-x.

8. Vansandt LM, Meinsohn MC, Godin P, et al. Durable contraception in the female domestic cat using viral-vectored delivery of a feline anti-Mullerian hormone transgene. Nat Commun. 2023;14(1):3140. PMID: 37280258. DOI: 10.1038/s41467-023-38721-0.

9. Ciccarelli M, Giassetti MI, Miao D, et al. Donor-derived spermatogenesis following stem cell transplantation in sterile NANOS2 knockout males. Proc Natl Acad Sci U S A. 2020;117(39):24195-24204. PMID: 32929012. DOI: 10.1073/pnas.2010102117.

10. Reese S, Pirez MC, Steele H, Kolle S. The reproductive success of bovine sperm after sex-sorting: a meta-analysis. Sci Rep. 2021;11(1):17366. PMID: 34462506. DOI: 10.1038/s41598-021-96834-2.

11. Grunwald HA, Gantz VM, Poplawski G, et al. Super-Mendelian inheritance mediated by CRISPR-Cas9 in the female mouse germline. Nature. 2019;566(7742):105-109. PMID: 30675057. DOI: 10.1038/s41586-019-0875-2.

Somatic Mosaicism Complicates Early Biomarker Validation: A Case for Longitudinal, Within-Person Monitoring

Hongryeol Park[1]

1Max Planck Institute for Molecular Biomedicine, Röntgenstraße 20, 48149 Münster Germany
Hongryeol.park@mpi-muenster.mpg.de

Abstract. Somatic mosaicism presents a fundamental challenge for early biomarker validation because the detection of a somatic mutation or selected clone does not necessarily indicate disease. Small clone size, spatial restriction, tissue heterogeneity, variable biomarker release and substantial inter-individual biological variation further limit the performance of cross-sectional approaches. This Commentary argues that longitudinal, within-person monitoring could complement conventional population-based biomarker strategies by identifying persistent and coordinated deviations from an individual's baseline. Integrating clonal dynamics with circulating molecular, immune and physiological measurements may help distinguish meaningful biological transitions from transient variation. Such temporal signals should then guide targeted spatial investigation and mechanistic validation. Combining population studies, personal longitudinal trajectories and experimental mosaic models may therefore provide a more robust framework for identifying context-dependent molecular changes that precede clinically apparent disease.

Keywords: Somatic mosaicism; Early detection; Longitudinal monitoring; Personal baselines

1. Introduction

Early detection is most straightforward when a disease signal is stable, widely distributed and clearly separated from normal variation. Somatically acquired changes rarely meet all three conditions. They arise in individual cells, expand at different rates and may remain confined to one tissue or cell population. Their clinical meaning also depends on when they appear, where they are located and how they interact with the surrounding tissue. These features make somatically driven diseases fundamentally different from disorders caused by a systemic inherited variant.

This distinction helps explain why successful germline screening cannot be transferred directly to somatic disease. An inherited pathogenic variant, such as a variant in BRCA1 or BRCA2, is present throughout the body and can be assessed from a single blood sample. A somatic variant may be present in only a small fraction of cells, may not enter the circulation and may never cause disease. Early detection therefore depends on more than analytical sensitivity. It also requires information about clone size, cell identity, anatomical location and biological context.

The study by Körber and colleagues illustrates both the power and the limits of detecting somatic selection. Their method, SCIFER, uses variant allele frequency (VAF) distributions from bulk whole-genome sequencing (WGS) to distinguish clonal selection from neutral drift. It can infer clone size, clone age and selective advantage without prior knowledge of the driver mutation. Applied to nonmalignant human bone marrow and brain, SCIFER identified widespread clonal selection, including selected clones with no known driver and brain clones carrying glioma-associated mutations [1]. However, detection still depended strongly on clone size and sequencing depth. At 90-fold coverage, clones at 5% VAF or above were detected reliably, whereas smaller clones were much harder to resolve [1].

These findings lead to the central argument of this Commentary. Somatic mutation, clonal selection and disease are closely related, but they represent distinct biological states. Somatic mosaicism is not the only reason that early biomarkers fail. However, it compounds existing limitations, including insufficient sensitivity, spatially restricted lesions, inconsistent release of biomarkers into the circulation, temporal variation and biases in study design. Improving early detection will therefore require methods that can more reliably detect small, localized and dynamically changing disease-associated signals.

2. Clonal Expansion Is Not the Same as Disease

An expanded clone is evidence that a cell population has gained a competitive advantage in a particular environment. It is not, by itself, evidence that the clone is pathogenic. Some selected clones can persist or occupy large areas of tissue for years without causing overt disease. Their fitness in normal tissue may also differ from their fitness during disease progression.

Normal human esophageal epithelium provides a striking example of how the effect of a mutation can depend on tissue context. By middle age, clones carrying inactivating NOTCH1 mutations can occupy a large proportion of otherwise healthy esophageal epithelium. Many of these clones carry biallelic alterations that disrupt NOTCH1 signaling. Lineage-tracing experiments in mice provided a functional explanation for this expansion: loss of one Notch1 allele increased the competitive fitness of esophageal progenitor cells, and loss of the second allele produced an even greater advantage.

The same loss of NOTCH1 function did not, however, promote tumor growth. NOTCH1 mutations were less common in esophageal tumors than in the surrounding normal epithelium, and genetic deletion or pharmacological blockade of Notch1 reduced the growth of established tumors in mice [2]. Thus, inactivation of NOTCH1 can promote clonal expansion during normal tissue maintenance while impairing cell proliferation after tumorigenesis has begun. This example shows that successful clonal expansion in normal tissue does not necessarily indicate that the same clone will promote disease.

A similar interpretive problem arises outside cancer. Somatic mutations involving STAT3, SOCS1, DNMT3A and related genes have been found in peripheral CD8+ T cells from people with multiple sclerosis. However, mutations were also detected in healthy controls, and their prevalence did not consistently distinguish the two groups [3]. These findings justify further investigation of mutated immune-cell clones, but they do not establish such clones as a cause or diagnostic marker of multiple sclerosis.

The diagnostic question is therefore not simply whether a mutation is present. It is whether a particular clone, in a particular cell type and tissue, is moving toward a pathogenic state. That judgment requires information about clone size, anatomical niche, co-occurring alterations, environmental pressures and change over time. An expanded clone may indicate selection without indicating disease.

3. Why Single-Timepoint Biomarkers Are Vulnerable

Most biomarker studies compare one sample from people with established disease with one sample from nominally healthy controls. This design can identify differences that accompany disease, but it is poorly suited to observing the transition into disease. At an early stage, the relevant clone may be small, spatially restricted or physiologically silent. A blood sample can therefore be negative even when an abnormal clone is already present elsewhere in the body.

Spatial restriction is a central part of this problem. Clonal hematopoiesis can often be sampled in blood because the affected cells circulate. By contrast, clonal expansions in the brain or in epithelial tissues such as the esophagus, colon and skin may remain local. Even a large tissue clone may release little or no measurable material into the circulation. A negative liquid-biopsy result may therefore reflect limited access to the lesion rather than the absence of disease-related biology.

Spatial heterogeneity also limits what can be inferred from a tissue biopsy. Sequencing a resected lesion can define the clone that was sampled, but it cannot be assumed to describe unsampled regions, separate foci or metastases. Multiregion sequencing of renal-cell carcinomas found that 63–69% of somatic mutations were not present in every tumor region, so a single biopsy captured only part of the genomic landscape [19]. Multifocal prostate cancers likewise showed marked differences among foci and, in some patients, independent clonal origins [20]. Thus, even direct access to one lesion does not reveal the complete somatic architecture of disease elsewhere in the same person.

This heterogeneity also affects treatment. A molecular target identified in one region may be absent from another or confined to only a subset of cells. Targeted therapy may therefore suppress sensitive clones while allowing pre-existing resistant clones to persist and expand. In colorectal cancers treated with epidermal growth factor receptor (EGFR) blockade, KRAS resistance mutations emerged during treatment, and modeling indicated that resistant subclones were already present before therapy began [21]. The implication is not that targeted therapy is generally ineffective, but that treatment decisions based on one sample may fail to account for all clinically relevant clones.

Circulating tumor DNA (ctDNA) illustrates the combined effects of abundance, location and shedding. In a study of several tumor types, ctDNA was detected frequently in advanced disease but less consistently in localized cancer. Detection rates for localized colorectal, gastroesophageal, pancreatic and breast cancers were approximately 73%, 57%, 48% and 50%, respectively [4]. Error-corrected sequencing has enabled direct detection of ctDNA in a proportion of stage I-II cancers, but sensitivity remains incomplete [5]. The limitation is not only technical: tumor size, vascular access, anatomical site and DNA release all influence whether a signal reaches the blood.

A positive plasma signal can also be misleading. Clonal hematopoiesis contributes many variants detected in circulating cell-free DNA (cfDNA), which can make a hematopoietic variant appear tumor-derived. Matched white-blood-cell sequencing is therefore important for identifying the source of plasma variants [6]. Together, these examples show why early biomarker validation is fragile. Low VAF, pre-analytical variation, analyte degradation, tissue-of-origin ambiguity, spatial heterogeneity, spectrum bias, lead-time bias and overdiagnosis can all reduce performance. Somatic mosaicism does not create every one of these problems, but it intensifies them by adding low-frequency, tissue-specific and clonally heterogeneous signals to the background. These interacting technical, biological and study-design barriers are summarized in Fig. 1.

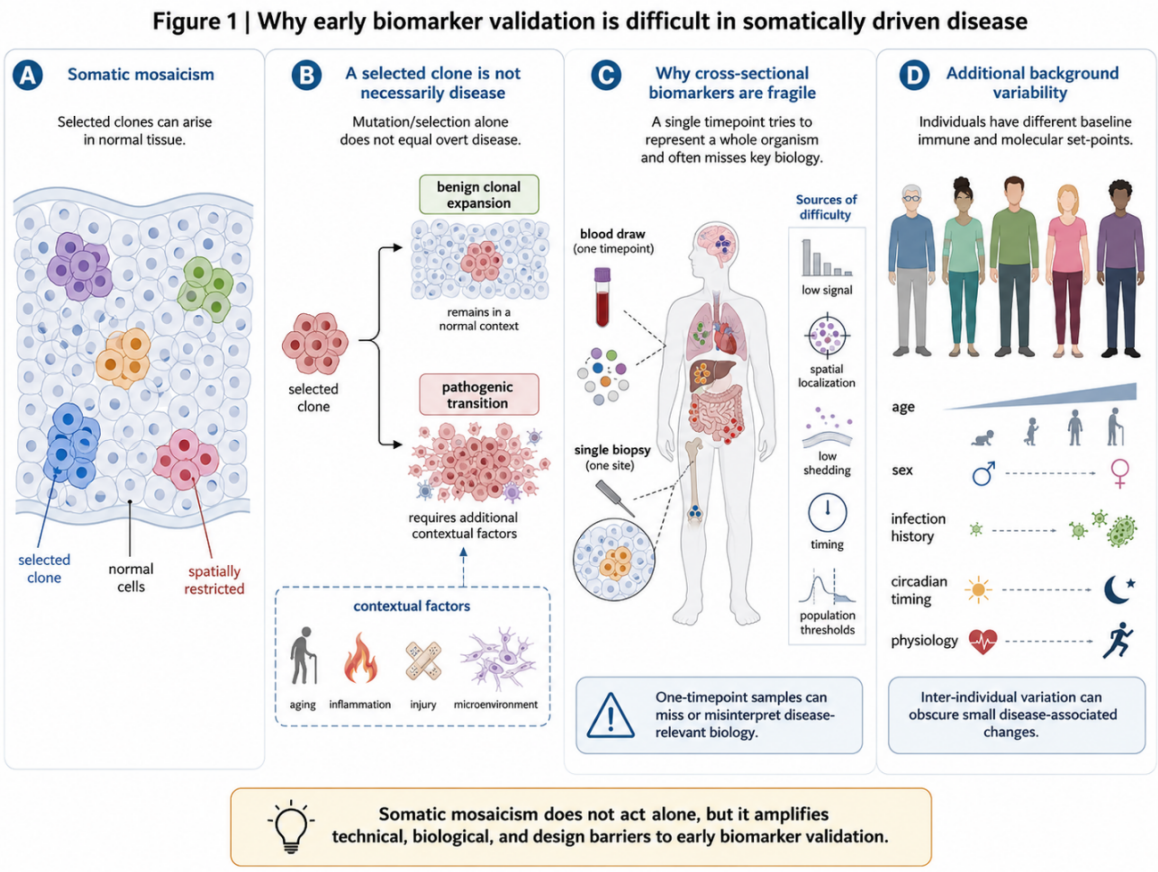


Fig. 1. Why early biomarker validation is difficult in somatically driven disease. Somatic mosaicism amplifies, but does not independently explain, technical, biological and study-design barriers.

4. Normal Biology Creates a Variable Background

Even when no disease-associated clone is present, healthy individuals differ markedly in their immune and molecular profiles. Repeated lymphocyte measurements are usually more similar within the same person than between different people, suggesting that each person maintains a partly characteristic immune-cell profile [7]. A one-year study of 99 healthy adults likewise found that the degree of longitudinal immune variation was itself an individual feature [8]. In a two-year multi-omic cohort, variation between individuals exceeded baseline variation within individuals across several molecular measurements, while plasma protein profiles were especially stable and person-specific [9].

These personal baselines are not fixed independently of context. The 10,000 Immunomes Project showed that leukocyte and cytokine distributions vary with age, sex and self-reported race [10]. Environmental history can have equally strong effects. In a twin study, most measured immune traits were dominated by non-heritable influences, and discordant cytomegalovirus (CMV) infection affected 119 of 204 immune parameters [11]. CMV seropositivity is associated with expansion of the CD8+ memory compartment and depletion of naive CD8+ T cells [12]. Circulating lymphocyte populations also vary across the day, making sampling time another source of biological variation [13].

These observations do not show that somatic mosaicism creates individual immune set-points. They show that any mosaic signal must be interpreted against a background shaped by genetics, age, sex, infection history, physiology and sampling conditions. A small early disease signal may be unremarkable relative to a broad population distribution yet represent a meaningful change for one person. This is a major limitation of static population thresholds.

5. What Longitudinal Monitoring Could Add

Longitudinal, within-person monitoring offers a way to reduce this uncertainty. Instead of asking only whether a result falls outside a population reference range, it asks whether the result has departed from that individual's expected trajectory. Each participant therefore serves as an internal reference. For measurements with demonstrated high individuality, including major T-lymphocyte subsets, reference-change values (RCVs) may complement population reference intervals [14].

The aim is not to reconstruct the complete mosaic architecture of the body from a single circulating analyte. That is unlikely to be possible. The more realistic goal is to follow several partial signals together: changes in cfDNA VAFs, methylation or fragmentation; shifts in circulating microRNAs (miRNAs); immune-cell composition; inflammatory proteins; metabolism; and physiological measurements. Each signal may be nonspecific in isolation. Their order, persistence and co-occurrence may be more informative than any single value.

A sustained increase in a clone-associated VAF becomes substantially more informative when it is followed by concordant changes in inflammatory proteins, immune-cell composition or organ-related physiology. The value of longitudinal monitoring therefore lies not in any single measurement, but in detecting persistent and temporally ordered changes across several biological layers. Such coordinated trajectories can distinguish a developing biological transition from short-term fluctuation and identify the point at which targeted investigation should begin. In this framework, disease risk emerges from the interaction between an expanding clone, its local tissue environment and external pressures rather than from the presence of a mutation alone.

Longitudinal monitoring should therefore be viewed as the first step in a staged diagnostic strategy. Its role is to determine when a potentially important transition is occurring, even before its anatomical source is known. Once a persistent coordinated pattern is detected, tissue-of-origin cfDNA features, imaging, organ-specific measurements or targeted tissue sampling can be used to determine where the change originates. Spatially resolved tissue analysis and mechanistic models can then establish whether the detected somatic clone is linked to local pathology and contributes to disease progression.

6. A Testable Implementation Framework

The following framework is a research proposal, not an established clinical schedule. A prospective study could enroll clinically healthy people at elevated risk and follow them for at least two years. The first objective would be methodological: determine whether within-person trajectories predict independently validated clinical or imaging-defined transitions better than cross-sectional thresholds.

Sampling strategy

A short run-in period should first define analytical and biological baseline variation. Three or four samples collected over six to eight weeks could be obtained under standardized conditions, including a similar time of day, fasting status and processing interval. Routine sampling could then occur quarterly in lower-risk participants and monthly in higher-risk groups. Additional samples could be collected after major events such

as infection, vaccination, surgery, pregnancy, menopause or initiation of an immune-modifying treatment. These intervals are illustrative and would need empirical optimization.

Sampling frequency should not be based on a universal age calendar. Plasma proteomic analyses have identified nonlinear waves near 34, 60 and 78 years, whereas a separate longitudinal multi-omic study found prominent changes around 44 and 60 years [15,16]. The difference probably reflects cohort composition, analytical platform and statistical method. Age and major life transitions may therefore help adjust sampling intensity, but current evidence does not justify fixed testing windows for the general population.

Panel architecture

The measurement platform should combine a stable core panel with separately validated exploratory modules. The core should directly measure clonal dynamics and the biological changes needed to interpret them. It could include:

cfDNA sequencing for selected low-VAF variants, methylation patterns, fragmentation features and broad copy-number changes; complete blood counts and reproducible immune-cell phenotyping; selected inflammatory, complement, coagulation and tissue-injury proteins; routine clinical chemistry; and contextual information on symptoms, medications, infections and relevant physiological changes.

Additional modalities, including circulating miRNA or broader small-RNA profiling, could be evaluated as exploratory modules. They should be added to the longitudinal panel only after demonstrating analytical stability and an interpretable relationship with clone-associated or disease-related changes.

When tissue information is available, variant selection should not rely on a single lesion-specific mutation. The panel should prioritize alterations shared across sampled regions and, where necessary, track several region-specific variants rather than treating one biopsy as representative of the whole disease.

Beyond selecting biologically informative measurements, a longitudinal study must also maintain analytical continuity over time. Changes in platform, reagents or calibration can create artificial trends that resemble biological change. Core measurements should therefore remain locked during a defined study phase. New markers or modules should be introduced only after analytical bridging using stored samples, reference materials and cross-platform calibration.

Analysis and validation

The first analytical task is to define each participant's expected range and identify changes that persist across repeated measurements. A single abnormal value is unlikely to be informative on its own. The analysis should instead look for coordinated changes in clone-associated VAFs, immune-cell composition, circulating proteins and physiological measurements. Reference-change values or change-point approaches may help identify these transitions, but the main aim is to distinguish sustained biological change from short-term fluctuation and measurement noise.

A coordinated circulating signal would not, by itself, reveal where the underlying process is occurring or prove that an expanding somatic clone is pathogenic. Its main value would be to identify when more targeted investigation is warranted. Imaging, tissue-of-origin cfDNA features and, when clinically justified and feasible, sampling of more than one suspicious site could then be used to locate the source of the change and determine whether it is associated with local tissue pathology. No sampled lesion should be assumed to represent all unsampled clones or disease sites.

Candidate patterns should subsequently be tested in independent longitudinal cohorts and, where possible, in spatially resolved human samples or experimental mosaic models. In this framework, longitudinal monitoring identifies when a potentially important transition occurs, targeted investigation determines where it originates, and mechanistic studies assess whether and how the somatic clone contributes to disease (Fig. 2).

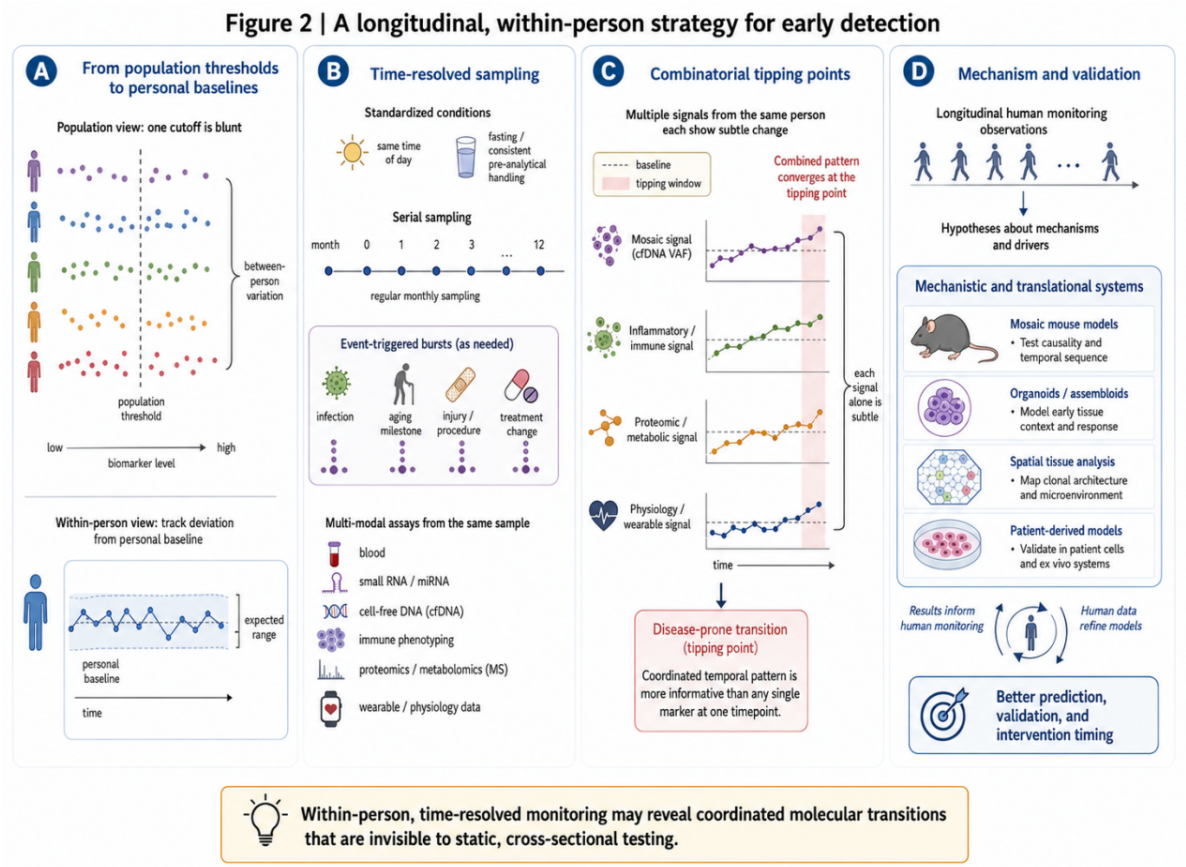


Fig. 2. A longitudinal, within-person strategy for early detection. Serial multimodal measurements are evaluated against personal baselines, and coordinated temporal transitions are subjected to spatial, mechanistic and translational validation.

7. Mechanistic Models Are Still Needed

Human longitudinal studies can identify candidate transitions, but they cannot explain why a clone becomes pathogenic. Controlled mosaic models are needed to test causality. Multicolour Confetti lineage tracing has enabled direct visualization of clonal competition, while the iFlpMosaics toolkit allows temporally controlled, multispectral comparison of mutant and wild-type cells within the same tissue environment [17,18]. These systems can test whether aging, inflammation, injury or a second mutation changes the behaviour of a pre-existing clone.

Mouse studies should be integrated with human organoids, spatially resolved tissue analysis and patient-derived models. The aim is not simply to reproduce a circulating biomarker signature. It is to determine whether the combinations identified in human time-series data alter clonal fitness, tissue homeostasis or disease initiation. This mechanistic step is essential before a temporal association can be interpreted as a pathogenic transition.

8. Conclusion

Somatic mosaicism does not make early biomarker discovery impossible, and it is not the only reason that candidate biomarkers fail. Its importance lies in making the relationship between molecular detection and disease less direct. A selected clone can remain benign, a disease-associated mutation can occur in a healthy person, a clinically important lesion can remain invisible in blood and a mutation found in one lesion may not represent other lesions in the same individual. The meaning of a signal therefore depends on tissue, cell type, clonal distribution, time and context.

The most defensible strategy is to combine population studies, longitudinal within-person monitoring and mechanistic validation. Population cohorts are needed to define risk, specificity and generalizability. Personal trajectories may reveal small but coordinated departures from an individual baseline. Mosaic experimental

systems are needed to determine whether those transitions are causal. Together, these approaches could move early detection away from the search for one universal static marker and toward the identification of context-dependent molecular transitions that precede clinically apparent disease.

Abbreviations

10KIP, 10,000 Immunomes Project; cfDNA, cell-free DNA; CMV, cytomegalovirus; ctDNA, circulating tumor DNA; EGFR, epidermal growth factor receptor; miRNA, microRNA; RCV, reference-change value; SCIFER, clonal-selection inference method introduced by Körber et al.; VAF, variant allele frequency; WGS, whole-genome sequencing.

References

1. Körber, V., Jakobsen, N. A., Ansari-Pour, N. et al. Detecting and quantifying clonal selection in somatic stem cells. Nature Genetics 57, 1718–1729 (2025). doi:10.1038/s41588-025-02217-y.
2. Abby, E., Dentro, S. C., Hall, M. W. J. et al. Notch1 mutations drive clonal expansion in normal esophageal epithelium but impair tumor growth. Nature Genetics 55, 232–245 (2023). doi:10.1038/s41588-022-01280-z.
3. Valori, M., Jansson, L. &Tienari, P. J. CD8+ cell somatic mutations in multiple sclerosis patients and controls—enrichment of mutations in STAT3 and other genes implicated in hematological malignancies. PLOS ONE 16, e0261002 (2021). doi:10.1371/journal.pone.0261002.
4. Bettegowda, C., Sausen, M., Leary, R. J. et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Science Translational Medicine 6, 224ra24 (2014). doi:10.1126/scitranslmed.3007094.
5. Phallen, J., Sausen, M., Adleff, V. et al. Direct detection of early-stage cancers using circulating tumor DNA. Science Translational Medicine 9, eaan2415 (2017). doi:10.1126/scitranslmed.aan2415.
6. Razavi, P., Li, B. T., Brown, D. N. et al. High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. Nature Medicine 25, 1928–1937 (2019). doi:10.1038/s41591-019-0652-7.
7. Sekiguchi, D. R., Smith, S. B., Sutter, J. A. et al. Circulating lymphocyte subsets in normal adults are variable and can be clustered into subgroups. Cytometry Part B: Clinical Cytometry 80, 291–299 (2011). doi:10.1002/cyto.b.20594.
8. Lakshmikanth, T., Muhammad, S. A., Olin, A. et al. Human immune system variation during 1 year. Cell Reports 32, 107923 (2020). doi:10.1016/j.celrep.2020.107923.
9. Tebani, A., Gummesson, A., Zhong, W. et al. Integration of molecular profiles in a longitudinal wellness profiling cohort. Nature Communications 11, 4487 (2020). doi:10.1038/s41467-020-18148-7.
10. Zalocusky, K. A., Kan, M. J., Hu, Z. et al. The 10,000 Immunomes Project: building a resource for human immunology. Cell Reports 25, 513–522.e3 (2018). doi:10.1016/j.celrep.2018.09.021.
11. Brodin, P., Jojic, V., Gao, T. et al. Variation in the human immune system is largely driven by non-heritable influences. Cell 160, 37–47 (2015). doi:10.1016/j.cell.2014.12.020.
12. Chidrawar, S., Khan, N., Wei, W. et al. Cytomegalovirus-seropositivity has a profound influence on the magnitude of major lymphoid subsets within healthy individuals. Clinical & Experimental Immunology 155, 423–432 (2009). doi:10.1111/j.1365-2249.2008.03785.x.
13. Ritchie, A. W., Oswald, I., Micklem, H. S. et al. Circadian variation of lymphocyte subpopulations: a study with monoclonal antibodies. British Medical Journal (Clinical Research Edition) 286, 1773–1775 (1983). doi:10.1136/bmj.286.6380.1773.
14. Falay, M., Senes, M., Korkmaz, S. et al. Biological variation of peripheral blood T-lymphocytes. Journal of Immunological Methods 470, 1–5 (2019). doi:10.1016/j.jim.2019.04.002.
15. Lehallier, B., Gate, D., Schaum, N. et al. Undulating changes in human plasma proteome profiles across the lifespan. Nature Medicine 25, 1843–1850 (2019). doi:10.1038/s41591-019-0673-2.
16. Shen, X., Wang, C., Zhou, X. et al. Nonlinear dynamics of multi-omics profiles during human aging. Nature Aging 4, 1619–1634 (2024). doi:10.1038/s43587-024-00692-2.
17. Snippert, H. J., van der Flier, L. G., Sato, T. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell 143, 134–144 (2010). doi:10.1016/j.cell.2010.09.016.
18. Garcia-Gonzalez, I., Gambera, S., Rocha, S. F. et al. iFlpMosaics enable the multispectral barcoding and high-throughput comparative analysis of mutant and wild-type cells. Nature Methods 22, 323–334 (2025). doi:10.1038/s41592-024-02534-w.
19. Gerlinger, M., Rowan, A. J., Horswell, S. et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. New England Journal of Medicine 366, 883–892 (2012). doi:10.1056/NEJMoa1113205.
20. Boutros, P. C., Fraser, M., Harding, N. J. et al. Spatial genomic heterogeneity within localized, multifocal prostate cancer. Nature Genetics 47, 736–745 (2015). doi:10.1038/ng.3315.
21. Diaz, L. A. Jr, Williams, R. T., Wu, J. et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature 486, 537–540 (2012). doi:10.1038/nature11219

Unresolved links between ammonia stress resilience and biomass production in microalgae grown in piggery wastewater

Eun Bin Joa[1] and Sang-Ah Lee[1]*

[1] Faculty of Biotechnology, College of Applied Life Sciences, Jeju National University, 102 Jejudaehak-ro, Jeju, 63243, Republic of Korea

leesa@jejunu.ac.kr

Abstract. Piggery wastewater (PWW) has attracted considerable interest as a nutrient rich medium for microalgal cultivation because it contains high concentrations of bioavailable nitrogen, organic carbon, phosphorus, and essential micronutrients. The biological conversion of these nutrients into algal biomass offers a promising approach to wastewater remediation and resource recovery while promoting circular carbon and nutrient use and reducing reliance on fossil derived resources. However, high total ammoniacal nitrogen in PWW can generate inhibitory concentrations of free ammonia (NH_3), particularly at elevated pH and temperature, reducing microalgal growth, biomass productivity, and treatment efficiency. As microalgae consume dissolved CO_2 during photosynthesis, the culture pH rises, shifting the NH_4^+/NH_3 equilibrium toward free ammonia and increasing cytotoxicity. Nevertheless, some microalgae have shown better biomass in PWW than would be predicted from ammonia concentration alone. It may reflect strain specific tolerance, stress modifying effects of wastewater matrix components, or both. However, the mechanisms underlying these effects remain poorly understood. This mini review critically evaluates molecular and metabolic responses translate into algal biomass production in PWW.

Keywords: Piggery wastewater, microalgae, free ammonia, stress resilience, organic carbon, transcriptomics

1. Introduction

With global population growth and the accompanying increase in demand for pork, pig farming has expanded substantially, generating large volumes of piggery wastewater (PWW). It contains high concentrations of organic matter, ammonia nitrogen, phosphorus, suspended solids, and micronutrients. Inadequate treatment and discharge of PWW can cause oxygen depletion, eutrophication, cytotoxicity to organisms, and deterioration of water quality. Conversely, its high nutrient content makes PWW a potentially valuable cultivation medium for microalgae, which can convert dissolved carbon and nutrients into recoverable biomass while reducing the pollutant load (Wang et al., 2015; Chen et al., 2020).

Several microalgae, particularly strains belonging to *Chlorella*, *Coelastrella*, and *Acutodesmus*, have demonstrated the ability to grow in PWW. Indigenous *Chlorella vulgaris* MBFJNU-1 removed more than 90% of total nitrogen and phosphorus from undiluted swine slurry (Wen et al., 2017), while *Chlorella sorokiniana* AK-1 achieved high biomass production and nutrient removal when cultivated in raw PWW, particularly after cell immobilization (Chen et al., 2020). Sequential application of ammonia adapted microalgae reduced ammonia from 850 to 81 mg L^{-1} in undiluted raw PWW within six days, although this process also used nitrogen starved, high density inocula (Lee et al., 2021b). These results demonstrate the feasibility of microalgae based PWW treatment but also reveal considerable variation among strains and process configurations. A quantitative strain comparison further illustrates this variation: the PWW derived *C. vulgaris* MBFJNU-1 tolerated 115 mg L^{-1} free ammonia, approximately three times the 38 mg L^{-1} tolerated by the reference strain UETX395 (Dai et al., 2023). Accordingly, performance under high ammonia conditions may reflect strain level tolerance as well as interactions with wastewater matrix components. Volatile fatty acids (VFAs), including acetate, propionate, and butyrate, can support mixotrophic growth and nutrient assimilation (Hu et al., 2012). Balancing nitrogen rich PWW with carbon rich brewery wastewater improved *C. vulgaris* biomass production and ammonia removal by correcting the carbon to nitrogen ratio (Zheng et al., 2018). However, studies integrating algal biomass bioconversion, wastewater treatment, and transcriptomic responses to ammonia stress remain scarce in Figure 1 (Priem et al., 2022).

This review evaluates the physicochemical and metabolic mechanisms underlying strain specific ammonia tolerance and matrix mediated stress alleviation and synthesizes transcriptomic evidence to determine whether these responses are linked to improved biomass productivity and wastewater treatment efficiency.

2. Freeammonia inhibition of microalgal piggery wastewater treatment bioprocess

Total ammoniacal nitrogen comprises ionized ammonium (NH_4^+) and unionized free ammonia (NH_3). Ammonium is an energetically favorable nitrogen source because it can be assimilated directly into glutamine and glutamate without prior nitrate reduction. However, high total ammoniacal nitrogen in PWW inhibit microalgal growth and reduce wastewater treatment efficiency. The fraction of unionized free ammonia increases with pH and temperature. Accordingly, microalgal growth correlates more closely with free ammonia concentration than with total ammonia nitrogen alone (Jiang et al., 2021; Zheng et al., 2019). During microalgal cultivation, photosynthetic CO_2 consumption raises culture pH, which shifts the NH_4^+/NH_3 equilibrium toward cytotoxic NH_3. Free ammonia leads to disruption of intracellular pH homeostasis and transmembrane proton gradients. It also damages to the oxygen evolving complex of photosystem II, and restriction of electron transfer from Q_a to Q_β (Wang et al., 2019). These disruptions induce reactive oxygen species formation, DNA damage, cell cycle checkpoint activation, and programmed cell death signaling (Chen et al., 2022). Finally, free ammonia exposure suppresses photosynthetic reactions, the glutamine synthetase, glutamate synthase, glutamate dehydrogenase pathways, ATP availability, reducing power, and carbon skeletons required for nitrogen assimilation in Figure 2 (Chen et al., 2022).

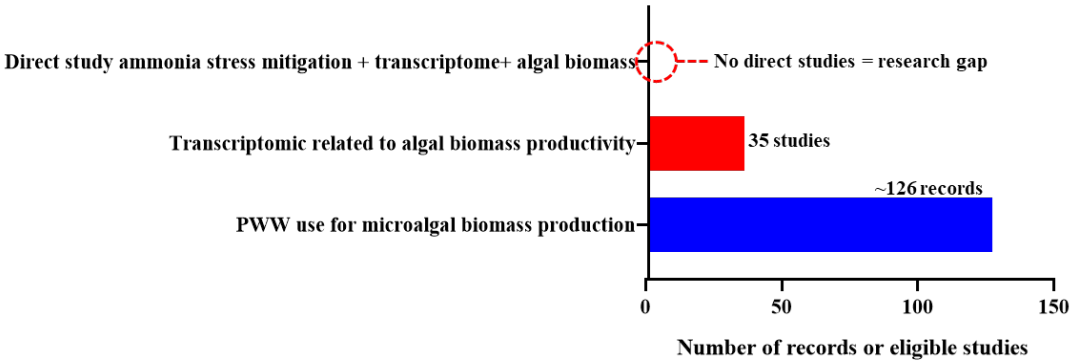


Fig. 1. Numbers of records and eligible studies across three research categories identified in OpenAlex using a publication-date filter covering 2006–2022.

3. Strain specific and matrix mediated mechanisms of ammonia stress resilience

3.1. Strain specific ammonia tolerance

Strain specific tolerance provides a distinct explanation for why some microalgae remain viable under high ammonia PWW conditions. The PWW derived *C. vulgaris* MBFJNU-1 tolerated 115 mg L^{-1} free ammonia, approximately three times the 38 mg L^{-1} tolerated by UETX395. Microalga coordinated arginine biosynthesis, photosynthetic electron generation/transfer, chloroplastic ATP synthesis, and glutathione metabolism (Dai et al., 2023). Similarly, the high ammonium tolerant *C. sorokiniana* tolerated 1000 mg L^{-1} NH_4^+ and downregulated multiple transporter genes, while maintaining intracellular homeostasis and assimilation (Xu et al., 2025). These findings suggest that strain level resilience involves controlled nitrogen entry together with sustained energy generation, nitrogen assimilation, and antioxidant protection. Nevertheless, NH_4^+ tolerance cannot be equated directly with free ammonia tolerance unless pH and temperature are reported, because the NH_3 fraction varies with both variables.

3.2. Matrix derived and exogenous organic carbon as mitigator of ammonia stress

3.2.1. acetate: the leading candidate in PWW

Acetate was detected at approximately 20 mM in the raw PWW examined by Lee et al. (2021a). It can be converted to acetyl CoA that supports the tricarboxylic acid and glyoxylate cycles, providing carbon for mixotrophic bioprocess. In C. vulgaris, acetate supplementation enhanced carbon utilization and energy metabolism compared with photoautotrophic cultivation (Yan et al., 2024). More recently, the introduction of acetatebased recovery stage following ammonium treatmentmaintainedculture pH near neutrality, decreased ammonia stripping by 41%, and enhanced biomass productivity up to 92% (Sun et al., 2026). The concurrent recovery of Fv/Fm suggested that acetate also contributed to restoration of photosystem II activity (Sun et al., 2026).

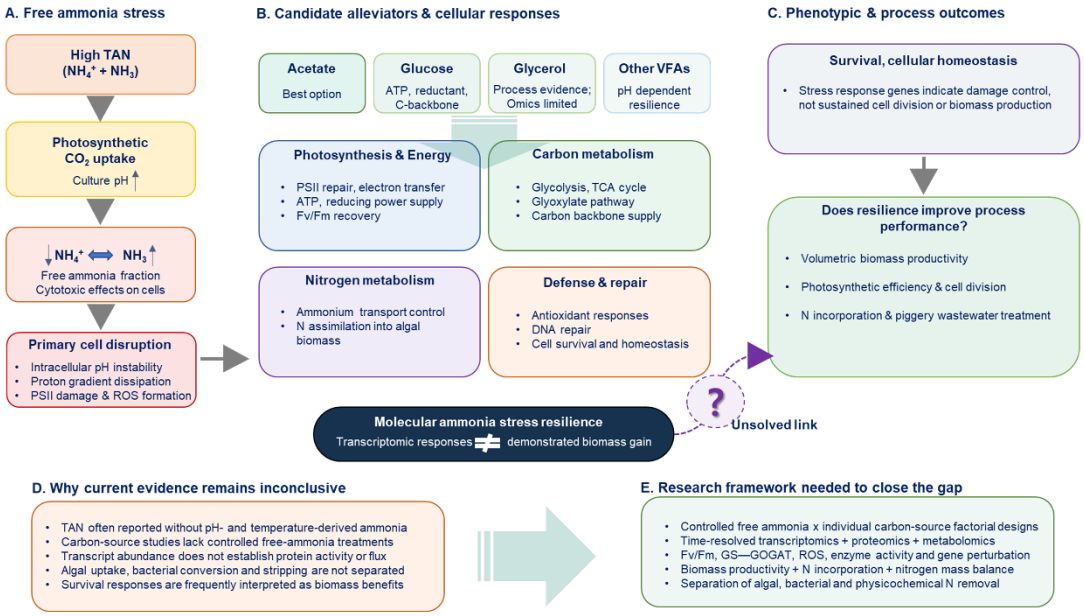


Fig. 2. Conceptual framework linking freeammonia stress responses to biomass production in microalgae grown in piggery wastewater (PWW). Photosynthetic CO_2 uptake increases culture pH and shifts the NH_4^+/NH_3 equilibrium toward cytotoxic free ammonia, disrupting cellular pH regulation, photosynthetic energy production, carbon and nitrogen metabolism, and cellular defense. Organic carbon sources, including acetate, glucose, glycerol, propionate, and butyrate, may alleviate these effects by modifying culture chemistry and supporting carbon and energy metabolism. However, the causal relationship between molecular stress resilience and improved biomass productivity remains unresolved. Addressing this gap requires controlled freeammoniacarbon source experiments integrating multi-omics, functional gene validation, physiological measurements, biomass productivity, and nitrogen mass balances. Solid arrows indicate established or proposed interactions, whereas dashed arrows represent unresolved causal relationships. TAN, total ammoniacal nitrogen; PSII, photosystem II; GS–GOGAT, glutamine synthetase–glutamate synthase; GDH, glutamate dehydrogenase; ROS, reactive oxygen species.

3.2.2. Glucose, glycerol, and other VFAs

Glucose is not generally a major constituent of PWW and is considered as a mechanistic analogue for carbon supported recovery rather than as an endogenous matrix derived alleviator. In Chlorella sp., glucose promoted ammonia assimilation more effectively than citrate, while glucose assisted nitrogen starvation increased cell viability and ammonia removal under high ammonia conditions (Lu et al., 2018). Glucose metabolism generates ATP and reducing power. It provides carbon skeletons that may support ammonium assimilation for cellular repair. Transcriptomic analysis of C. vulgaris under mixotrophic cultivation showed changes in replication/repair genes, thioredoxins, and inositol-3-phosphate synthase associated with oxidative stress

tolerance (Vidotti et al., 2020). Glucose also altered carbon fixation, carotenoid biosynthesis, and lipid metabolism in C. sorokiniana (Azaman et al., 2020). However, these studies lacked controlled free ammonia stress conditions; thus, they offer mechanistic insights rather than direct evidence of stress alleviation in microalgae cultivated in PWW. Glycerol addition increased NH_4^+-N removal efficiency by 96% in C. vulgaris cultivated in PWW, while the culture medium C/N ratio and pH also influenced biomass formation and ammonia removal (Zheng et al., 2019). Propionate and butyrate might support mixotrophic growth, but their effects depend on substrate concentration and pH; maintaining pH 8 decreased the undissociated VFA fraction. Microalga can be grown at VFA concentrations up to 5 g C L^{-1}. The biomass yield declined at higher concentrations of VFA (Lacroux et al., 2021).

Therefore, organic carbon can support stress recovery at the physiological level by increasing the availability of ATP, reducing power, and carbon skeletons. These effects may be reflected transcriptionally in photosystem II repair, antioxidant defense, and nitrogen assimilation pathways. However, pathway regulation under inoculation of stress mitigator does not establish free ammonia specific alleviation unless the molecular response is compared at the same free ammonia exposure with and without the carbon source.

4. Limitations of transcriptomic studies in linking stress responses to biomass production

Transcriptomic analyses have identified pathways potentially involved in microalgal ammonia tolerance, including photosynthetic electron transport, ATP production, nitrogen assimilation, antioxidant defense, DNA repair, and ammonium transport (Chen et al., 2022; Dai et al., 2023; Xu et al., 2025). However, limited number of studies have explored transcriptomic responses to free ammonia. Studies combining free ammonia exposure or raw PWW with biomass productivity measurements remain scarce. Most transcriptomic studies of organic carbon supported cultivation have been conducted under general mixotrophic conditions without controlled free ammonia exposure. Consequently, the metabolic pathways changes may represent general growth enhancement rather than ammonia specific stress resilience. Hence, future studies should investigate comprehensive experimental designs combining free ammonia concentrations with individual carbon sources and integrate transcriptomics with photosynthetic efficiency, nitrogen mass balances, and microalgal biomass productivity.

Moreover, a stress responsive transcriptional profile does not necessarily translate into enhanced microalgal biomass production. Upregulation of key gene pathways may indicate damage control and survival rather than sustained cell division, while downregulation of ammonium transporters may limit toxicity but also restrict nitrogen supply for biomass synthesis. To establish a stress signaling cascade, it is essential to correlate gene expression with key physiological parameters, including volumetric biomass productivity, photosynthetic efficiency, cell division, and nitrogen incorporation into biomass. RT-qPCR can also validate the expression of key genes, whereas functional roles require gene disruption, knockdown, or overexpression together with physiological and enzyme activity measurements.

5. Conclusions

Microalgae represent a promising biological platform for converting nutrients in PWW into valuable biomass. However, free ammonia toxicity remains a major constraint on biomass productivity and treatment efficiency. Both strain specific tolerance mechanisms and organic carbon supported metabolism may improve stress resilience, but their relative contributions to biomass production remain unresolved. Controlled free ammonia experiments comparing tolerant and sensitive strains with and without PWW relevant carbon sources, integrated with multi omics, physiological measurements, and nitrogen mass balances, are required to establish this relationship.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS- 2024- 00358292)

References

1. Azaman, S.N.A., Wong, D.C.J., Tan, S.W., Yusoff, F.M., Nagao, N., Yeap, S.K., 2020. De novo transcriptome analysis of *Chlorella sorokiniana*: Effect of glucose assimilation and moderate light intensity. Sci. Rep. 10, 17331. https://doi.org/10.1038/s41598-020-74410-4.
2. Chen, C.-Y., Kuo, E.-W., Nagarajan, D., Ho, S.-H., Dong, C.-D., Lee, D.-J., Chang, J.-S., 2020. Cultivating *Chlorella sorokiniana* AK-1 with swine wastewater for simultaneous wastewater treatment and algal biomass production. Bioresour. Technol. 302, 122814. https://doi.org/10.1016/j.biortech.2020.122814.
3. Chen, Z., Qiu, S., Li, M., Zhou, D., Ge, S., 2022. Instant inhibition and subsequent self-adaptation of *Chlorella* sp. Toward free ammonia shock in wastewater: Physiological and genetic responses. Environ. Sci. Technol. 56, 9641–9650. https://doi.org/10.1021/acs.est.1c08001.
4. Dai, J., Zheng, M., He, Y., Zhou, Y., Wang, M., Chen, B., 2023. Real-time response counterattack strategy of tolerant microalgae *Chlorella vulgaris* MBFJNU-1 in original swine wastewater and free ammonia. Bioresour. Technol. 377, 128945. https://doi.org/10.1016/j.biortech.2023.128945.
5. Hu, B., Min, M., Zhou, W., Du, Z., Mohr, M., Chen, P., Zhu, J., Cheng, Y., Liu, Y., Ruan, R., 2012. Enhanced mixotrophic growth of microalga *Chlorella* sp. On pretreated swine manure for simultaneous biofuel feedstock production and nutrient removal. Bioresour. Technol. 126, 71–79. https://doi.org/10.1016/j.biortech.2012.09.031.
6. Jiang, R., Qin, L., Feng, S., Huang, D., Wang, Z., Zhu, S., 2021. The joint effect of ammonium and pH on the growth of *Chlorella vulgaris* and ammonium removal in artificial liquid digestate. Bioresour. Technol. 325, 124690. https://doi.org/10.1016/j.biortech.2021.124690.
7. Lacroux, J., Seira, J., Trably, E., Bernet, N., Steyer, J.-P., van Lis, R., 2021. Mixotrophic growth of *Chlorella sorokiniana* on acetate and butyrate: Interplay between substrate, C:N ratio and pH. Front. Microbiol. 12, 703614. https://doi.org/10.3389/fmicb.2021.703614.
8. Lee, S.-A., Ko, S.-R., Lee, N., Lee, J.-W., Le, V.V., Oh, H.-M., Ahn, C.-Y., 2021a. Two-step microalgal (*Coelastrella* sp.) treatment of raw piggery wastewater resulting in higher lipid and triacylglycerol levels for possible production of higher-quality biodiesel. Bioresour. Technol. 332, 125081. https://doi.org/10.1016/j.biortech.2021.125081.
9. Lee, S.-A., Lee, N., Oh, H.-M., Ahn, C.-Y., 2021b. Stepwise treatment of undiluted raw piggery wastewater using three microalgal species adapted to high ammonia. Chemosphere 263, 127934. https://doi.org/10.1016/j.chemosphere.2020.127934.
10. Lu, Q., Chen, P., Addy, M., Zhang, R., Deng, X., Ma, Y., Cheng, Y., Hussain, F., Chen, C., Liu, Y., Ruan, R., 2018. Carbon-dependent alleviation of ammonia toxicity for algae cultivation and associated mechanisms exploration. Bioresour. Technol. 249, 99–107. https://doi.org/10.1016/j.biortech.2017.09.175.
11. Priem, J., Piwowar, H., Orr, R., 2022. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv 2205.01833. https://doi.org/10.48550/arXiv.2205.01833.
12. Sun, Z., Bo, C., Cao, S., Sun, L., 2026. Acetate-enabled staged strategy in *Chlorella vulgaris* culture: Relieving ammonium inhibition while reducing ammonia stripping and enhancing protein content. Algal Res. 94, 104527. https://doi.org/10.1016/j.algal.2026.104527.
13. Vidotti, A.D.S., Riaño-Pachón, D.M., Mattiello, L., Giraldi, L.A., Winck, F.V., Franco, T.T., 2020. Analysis of autotrophic, mixotrophic and heterotrophic phenotypes in the microalgae *Chlorella vulgaris* using time-resolved proteomics and transcriptomics approaches. Algal Res. 51, 102060. https://doi.org/10.1016/j.algal.2020.102060.
14. Wang, J., Zhou, W., Chen, H., Zhan, J., He, C., Wang, Q., 2019. Ammonium nitrogen tolerant *Chlorella* strain screening and its damaging effects on photosynthesis. Front. Microbiol. 9, 3250. https://doi.org/10.3389/fmicb.2018.03250.
15. Wang, Y., Guo, W., Yen, H.-W., Ho, S.-H., Lo, Y.-C., Cheng, C.-L., Ren, N., Chang, J.-S., 2015. Cultivation of *Chlorella vulgaris* JSC-6 with swine wastewater for simultaneous nutrient/COD removal

and carbohydrate production. Bioresour. Technol. 198, 619–625. https://doi.org/10.1016/j.biortech.2015.09.067.

16. Wen, Y., He, Y., Ji, X., Li, S., Chen, L., Zhou, Y., Wang, M., Chen, B., 2017. Isolation of an indigenous *Chlorella vulgaris* from swine wastewater and characterization of its nutrient removal ability in undiluted sewage. Bioresour. Technol. 243, 247–253. https://doi.org/10.1016/j.biortech.2017.06.094.
17. Xu, L., Chen, L., Jiang, L., Zhang, J., Wu, P., Wang, W., 2025. *Chlorella*'s transport inhibition: A powerful defense against high ammonium stress. Ecotoxicol. Environ. Saf. 290, 117460. https://doi.org/10.1016/j.ecoenv.2024.117460.
18. Yan, X., Shan, S., Li, X., Xu, Q., Yan, X., Ruan, R., Cheng, P., 2024. Carbon and energy metabolism for the mixotrophic culture of *Chlorella vulgaris* using sodium acetate as a carbon source. Front. Microbiol. 15, 1436264. https://doi.org/10.3389/fmicb.2024.1436264.
19. Zheng, H., Liu, M., Lu, Q., Wu, X., Ma, Y., Cheng, Y., Addy, M.M., Liu, Y., Ruan, R., 2018. Balancing carbon/nitrogen ratio to improve nutrients removal and algal biomass production in piggery and brewery wastewaters. Bioresour. Technol. 249, 479–486. https://doi.org/10.1016/j.biortech.2017.10.057.
20. Zheng, H., Wu, X., Zou, G., Zhou, T., Liu, Y., Ruan, R., 2019. Cultivation of *Chlorella vulgaris* in manure-free piggery wastewater with high-strength ammonium for nutrient removal and biomass production: Effects of ammonium concentration, carbon/nitrogen ratio and pH. Bioresour. Technol. 273, 203–211. https://doi.org/10.1016/j.biortech.2018.11.019.

Environment and Energy Division

Aims

The Environment and Energy Division invites voluntary contributions from VeKNI members that reflect the broad and rapidly evolving research landscape across the life sciences. Members are encouraged to share recent scientific advances, summaries of ongoing research projects, emerging technology trends, translational applications, industrial case studies, or perspectives on future challenges in biological and biomedical research.

Scope

- **Climate, Environment & Earth Systems:** Climate science, atmospheric and environmental processes, Earth system dynamics, environmental change, air and water quality, geosciences, and environmental monitoring.

- **Energy & Sustainability:** Sustainable energy systems, energy transition, energy storage, renewable energy, decarbonization strategies, resource efficiency, and sustainability-driven innovations.

- **Natural Resources & Circular Economy:** Resource management, circular economy, sustainable materials, recycling and recovery, waste management, responsible production and consumption, and environmental stewardship.

- **Ecosystems, Water & Environmental Resilience:** Ecosystem services, biodiversity, hydrology, marine and terrestrial environments, nature-based solutions, environmental resilience, and adaptation to global change.

- **Food, Society & Sustainable Development:** Sustainable food systems, agriculture, environmental governance, policy and management, stakeholder engagement, socioeconomic dimensions of sustainability, and sustainable development goals.

- **Digital Transformation & Data-Driven Sustainability:** Artificial intelligence, environmental informatics, modeling and simulation, digital technologies, decision-support systems, and data-driven approaches for sustainability research and management.

Maintenance Challenges and Inspection Technologies for Fixed-Bottom Offshore Wind Substructures in Europe

Martin S. D. Yoo and Jens Cruse

CRUSE Offshore GmbH, Eppendorfer Marktplatz 10, 20251 Hamburg, Germany
martin.yoo@cruse-offshore.de

Abstract. This paper reviews the main maintenance challenges and inspection technologies for fixed-bottom offshore wind substructures in Europe. The discussion focuses on monopiles, which account for the dominant share of installed foundations, and on the main degradation mechanisms relevant to maintenance: biofouling, corrosion, fatigue and seabed scour. Conventional diver-based inspection remains important where cleaning, close visual examination and minor repair are required, but it is constrained by weather, safety and working time. Remotely operated vehicles (ROVs) are increasingly used for visual surveys, sonar-based surveys, thickness measurements and selected non-destructive testing (NDT) methods. Autonomous underwater vehicles (AUVs) offer longer-term potential, although their repair capability and reliable operation close to complex structures remain limited. Thispaper also considers positioning, sensing, current industrial practice, and relevant standards and European research programmesthat support maintenance planning and technology development. In the near term, priority should be given not to complete autonomy, but to the systematic improvement of ROV-based inspection, data quality and repeatability. Such an approach can reduce diver exposure, support condition-based maintenance and contribute to safer and more cost-effective maintenance ofoffshore wind farms.

Keywords: offshore wind, substructure, maintenance, inspection, ROV, AUV

1. Introduction

Europe had installed approximately 37 GW of offshore wind capacity by the end of 2024 [1]. Most of this capacity is located in the North Sea and Baltic Sea, where fixed-bottom turbines remain the prevailing solution. Monopiles represented about 81% of European offshore wind foundations in the 2020 statistics [2], and their share has remained high as fabrication methods have enabled larger diameters and heavier sections. The reliability of these foundations is therefore central to the availability and lifetime of the European offshore wind fleet.

Unlike onshore structures, offshore foundations are simultaneously exposed to wind loading, waves, currents, seawater and seabed processes. Biofouling may obscure coating damage, corrosion and surface-breaking cracks from visual detection. At welded joints and other stress-concentrated details, cyclic loading combined with corrosion can accelerate fatigue damage. Scour around the foundation changes the support conditions and may increase bending moments and fatigue damage. These mechanisms do not usually occur in isolation; their interaction makes inspection planning and interpretation difficult. The main environmental actions and the location of the substructure are illustrated in Fig. 1.

This paper condenses the findings of a broader review of European maintenance practice. It addresses three questions: which degradation mechanisms require the greatest attention, which inspection and maintenance methods are currently practical, and where automation can provide practical benefits. The paper is based on published studies, industrial information and European standards rather than on new experiments.

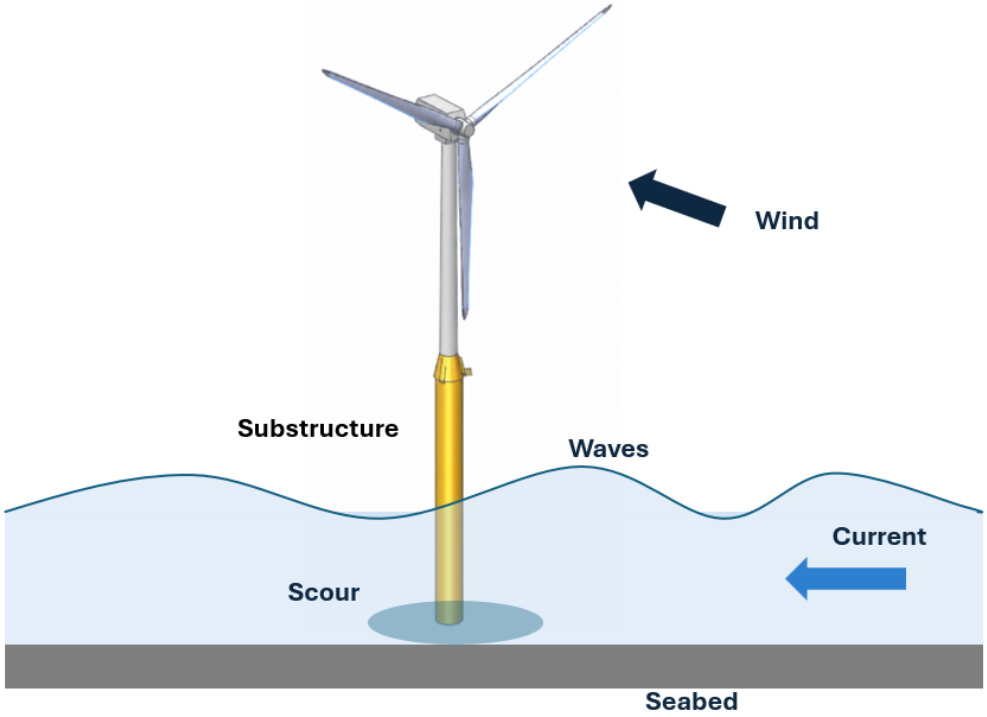


Fig. 1. Offshore wind turbine exposed to wind, waves, current and seabed scour.

2. Main Degradation Mechanisms and Protection

Biofouling begins within hours after immersion and develops from an initial film into microbial and larger biological communities. Its thickness and composition depend on water depth, nutrients, waves and sediment movement. Measurements in the Dutch North Sea indicate that depth and chlorophyll concentration are important variables [3]. Biofouling adds hydrodynamic roughness and, more importantly for maintenance, obscures the underlying foundation surface. Cleaning is therefore often required before a meaningful visual or nondestructive examination can bedone.

Corrosion conditions differ markedly between the atmospheric, splash, tidal, submerged and buried zones. The splash and tidal zones are particularly severe because repeated wetting and drying, oxygen supply and wave-induced mechanical effects promote rapid deterioration. In the submerged zone, local coating defects can promote localized corrosion and pitting, which may be difficult to detect visually. Corrosion fatigue is particularly importantin welded joints, where corrosion accelerates crack initiation and growth under cyclic loading. Reported corrosion rates for unprotected steelare generally higher in the splash zone than in the continuously submerged zone [4], although actual rates vary widely with environmental exposure and protection conditions.

Corrosion protection commonly combines protective coatings and cathodic protection, depending on the exposure zone and design concept. Thermal-sprayed aluminiumand high-build epoxy systems are widely used, with coating systems selected according to the exposure environment and required durability; ISO 12944-2 [5] provides the relevant environmental and corrosivity classifications.Galvanic anode cathodic protection is simple and passive, but sacrificial anodes must be sized for the required design life and may require replacement or supplementation if operation is extended beyond the original design period.Impressed-current systems use an external power supply and can be monitored and controlled, but they are more complex. Maintenance therefore includes coating surveys, potential measurements, anode condition checks and, where necessary, local coating repair or reinforcementof the cathodic-protection system.

Scour is the removal of seabed material by currents and waves around a foundation. It reduces effective embedded length and alters the soil–structure stiffness. The resulting increase in cyclic bending can shorten fatigue life. Scour depth depends on the wave–current regime, pile geometry and seabed characteristics; the classical work of Sumer and Fredsøe remains a useful basis for understanding these relationships [6]. Rock armour is the most common protection method, while mattresses and other bed-stabilisation concepts are also available. Scour protection does not eliminate the need for inspection, because displacement of armourstones and edge scour can develop over time.

3. Inspection, Repair and Automation

Diver inspection provides direct access to the structure and remains valuable for close visual inspection, cleaning, manual measurements and minor repair. Divers can perform ultrasonic thickness measurements and selected NDT methods such as magnetic particle inspection, although reliable results depend on adequate surface preparation, access and operator competence; visual methods are additionally limited by underwater visibility. Working time is limited, and operation risk increaseswith depth, strong currents and wave conditions. The need for specialised diving teams, support vessels and repeated offshore mobilisation can also make routine inspection campaigns costly.

Remotely operated vehicles (ROVs) reduce human exposure and allow longer inspection periods. A typical inspection ROV carries high-resolution cameras, lighting, sonar and navigation sensors. Larger work-class units can add cleaning brushes, water jets, manipulators and nondestructive-testing modules. In current European practice, ROVs are used for visual surveys, marine-growth assessment, cathodic-protection checks, wall- thickness measurement, weld inspection and multibeam surveys of scour and cable routes. ROV-based subsea inspections have, for example, been carried out at the Amrumbank West offshore wind farm to inspect monopiles, scour protection and cables [7].

Autonomous underwater vehicles (AUVs)remove the tether and can cover larger areas and potentially reduce dependence on support vessels. Their navigation commonly combines inertial sensors and Doppler velocity logs, often supplemented by acoustic positioning and depth measurements. However, acoustic communication is slow, battery capacity is finite, and reliable inspection close to a monopile or within geometrically complex structuresis more demanding than open-water surveying. Repair functions are still uncommon. AUVs are therefore better regarded as a complementary technology rather than an immediate replacement for ROVs or divers.

Near-term priorities forautomation includestable vehicle control, repeatable survey paths, accurate positioning and consistent data collection. Ultra-short baseline orlong baseline acoustic positioning, inertial navigation and Doppler velocity logs support accurate underwater navigation and positioning. Auto-hover and station-keeping functions help an ROV maintain position in strong currents, while sonar assists operation in poor visibility. Better repeatability allows measurements from successive inspections to be compared and supports condition-based maintenance. Fully autonomous defect recognition and repair are not a prerequisite for achieving substantial operational benefits; reliable sensor integration and consistent data management are more immediate priorities. Table 1 summarizes the practical strengths, limitations and near-term roles of divers, ROVs and AUVs in underwater inspection and maintenance.

Aspect	Diver	ROV	AUV
Control	Direct human operation	Real-time remote control through tether	Pre-programmed or supervised autonomy
Typical strengths	Cleaning, close inspection, measurements and minor repair	Repeatable surveys, sonar and selected NDT	Efficient wide-area and repeated surveys
Main limits	Safety, sea conditions and limitedworking time	Tether and support-vessel dependence; limited repair capability	Energy, communication and close-structure navigation
Near-term role	Selected close-contact and repair tasks	Main platform for remotely operated inspection	Wide-areasurveys and future resident inspection

Table 1. Practical comparison of underwater maintenance approaches.

4. European Industrial and Regulatory Context

Europe has a well-establishedmarine-service sector with experience from offshore oil and gas, subsea cables and wind farms. Companies such as Deutsche Windtechnik, CWind, TSC Subsea, JD-Contractor and Boskalis collectively provide a broad range of capabilities, including diving, ROV operations, bathymetric surveys, cathodic-protection inspection, cable inspection, subsea repair and seabed intervention. Available systems range from small observation ROVs to work-class vehicles equipped with manipulators and specialised inspection or NDT tools. This diversity is important because no single platform is technically and economically suited to all inspection and repair tasks.

There is no single European standard devoted exclusively to offshore wind maintenance. Instead, operators rely on a combination of IEC wind-turbine standards, ISO asset-management standards, national requirements, and DNV standards and recommended practices.DNV-ST-0126 concerns wind-turbine support structures, DNV-ST-0262 deals with lifetime extension, DNV-RP-C203 provides guidance on fatigue design, and DNV-RP-0416 addresses corrosion protection and in-service inspection.Together, these documents form a framework for structural integrity management, maintenance planning and decisions on continued operation.

European wind-energyresearch priorities increasingly emphasisedigitalisation, robotics and lifetime extension. The ETIPWind Strategic Research and Innovation Agenda for 2025-2027 identifies semi-automated inspection, advanced repair methods and autonomous marine operations as key research areas [8]. For industry, however, the value of automation will ultimately depend on reduced vessel time, lower personnel exposure, improved inspection quality and better information for maintenance decisions. Demonstration projects should therefore be evaluated primarily against such operational benefitsrather than the level of autonomy alone.

5. Conclusions

Offshore wind substructure maintenance is strongly influenced by biofouling, corrosion, fatigue and scour. Effective management requires suitable protection systems, periodic inspection and timely interventionwhere deterioration is detected.

Divers continue to play an important role in selected close-contact inspections and repair tasks, particularly where direct manipulation or close human observationis required. For many current underwater inspection tasks, ROVs provide a practical balance between technological maturity, sensor capability, endurance and reduced diver exposure. AUVs can support wider-area and repeated surveys, although operation close to complex structures remains challenging and repair capabilities are still limited.

A practical route toward greater automation is therefore likely to be gradual, with priorities including improved vehicle stability and positioning, repeatable survey paths, integration of cleaning and NDT tools, and consistent data collection over the service life of the structure. Rather than aiming immediately at complete autonomy, this stepwise approach addresses current operational needs and can contribute to safer, more reliable and more cost-effective maintenance of offshore wind farms.

References

1. WindEurope: Wind Energy in Europe: 2024 Statistics and the Outlook for 2025-2030. WindEurope, Brussels (2025).

2. WindEurope: Offshore Wind in Europe: Key Trends and Statistics 2020. WindEurope, Brussels (2021).

3. Almeida, L.P., Coolen, J.W.P.: Modelling thickness variations of macrofouling communities on offshore platforms in the Dutch North Sea. Journal of Sea Research 156 (2020).

4. Momber, A.W., et al.: Investigating corrosion protection of offshore wind towers. Journal of Protective Coatings & Linings, April (2008).

5. ISO 12944-2:2017: Paints and varnishes - Corrosion protection of steel structures by protective paint systems - Part 2: Classification of environments. ISO, Geneva (2017).

6. Sumer, B.M., Fredsoe, J.: The Mechanics of Scour in the Marine Environment. World Scientific, Singapore (2002).

7.En-former: Diving robot facilitates offshore wind farm inspections, https://www.en-former.com/en/diving-robot-facilitates-offshore-wind-farm-inspections/, last accessed 2025/06/07.

8. ETIPWind: Strategic Research & Innovation Agenda 2025-2027. ETIPWind (2023).

Highly Oxygenated Organic Molecules in Atmosphere: Formation, Detection, and Role in Secondary Organic Aerosol

Sungah Kang[1]

[1]Forschungszentrum Jülich GmbH, Germany

s.kang@fz-juelich.de

Abstract. Highly oxygenated organic molecules (HOM) are an important class of oxidation products in atmospheric chemistry. HOM are primarily formed through autoxidation, a process involving sequential intramolecular hydrogen shifts and oxygen (O_2) addition reactions following the initial oxidation of volatile organic compounds (VOCs). Autoxidation efficiently produces highly oxygenated organic peroxy radicals (HOM-RO_2•). These HOM-RO_2• radicals subsequently react with atmospheric reaction partners, such as other peroxy radicals (RO_2•), hydroperoxyl radicals (HO_2•), and nitric oxide (NO), leading to the formation of closed-shell HOM products. Many HOMhave low to extremely low vapor pressures, allowing them to efficiently condense onto existing particles and even contribute to new particle formation. Consequently, HOM are important contributors to the formation and growth of secondary organic aerosol (SOA). SOA constitutes a substantial fraction of atmospheric organic aerosol and can significantly influence climate and human health. Recent developments in advanced chemical ionization mass spectrometry have enabled molecular-level observations of HOM and provided new insights into their formation pathways and atmospheric chemistry. This paper briefly introduces the formation mechanisms and chemistry of HOM, analytical techniques used for their measurement, and their contribution to secondary organic aerosol formation.

Keywords: highly oxygenated organic molecules, HOM, autoxidation, peroxy radicals, VOC, secondary organic aerosol, atmospheric chemistry

1. Introduction

Peroxy radicals (RO_2) are important reactive intermediates in atmospheric chemistry and play a key role in the oxidation of volatile organic compounds (VOCs). VOCs are emitted from both biogenic sources, such as vegetation, and anthropogenic sources. Once emitted, VOCs undergo oxidation initiated by hydroxyl radicals (•OH), ozone (O_3), and nitrate radicals (NO_3•), producing oxygenated compounds and organic peroxy radicals. Further oxidation can decrease the volatility of these compounds, facilitating partitioning of gas phase compounds to particle phase and contributing to secondary organic aerosol (SOA) formation.

The identification of highly oxygenated organic molecules (HOM) has significantly improved our understanding of atmospheric organic oxidation and SOA formation. HOM contain multiple oxygen-containing functional groups and are mainly produced through a process known as autoxidation [1]. During autoxidation, rapidly increasing their oxygen content. The resulting highly oxygenated peroxy radicals (HOM-RO_2•) can react with other atmospheric radicals, such as RO_2• and HO_2•, or with nitric oxide (NO), forming closed-shell HOM products.

Due to their high oxygen content, many HOM have low to extremely low volatility, allowing them to efficiently condense onto existing aerosol particles and contribute to particle growth. Some HOM can also participate in the initial stages of new particle formation. Therefore, understanding HOM formation mechanisms and chemistry is crucial for improving our knowledge of SOA formation and its broader impacts on climate, air quality, and human health.

2. Formation of Highly Oxygenated Organic Molecules

VOCs can be oxidized by •OH, O_3, or NO_3•. After the initial oxidation step, alkyl radical is produced and oxygen addition happens extremely fast producing peroxy radical.

$R\bullet + O_2 \rightarrow RO_2\bullet$

This organic peroxy radical can undergo bimolecular reactions with $HO_2\bullet$, NO, or other $RO_2\bullet$ radicals. However, some $RO_2\bullet$ radicals undergo rapid intramolecular hydrogen shifts, producing alkyl radical site where oxygen is added resulting peroxy radical with two more oxygen atom in the molecule. This process is called autoxidation. It is rapid and effective process resulting in peroxy radical with high oxygen contents ($HOM\text{-}RO_2\bullet$)

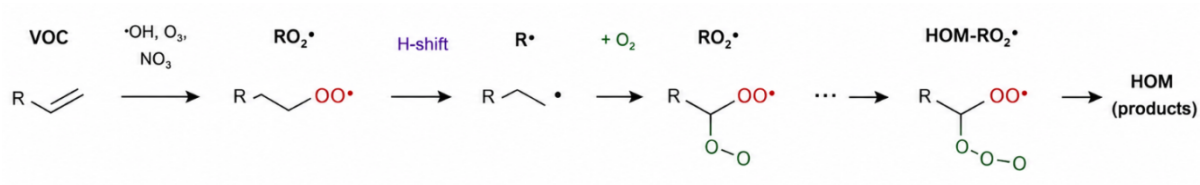


Fig. 1. Schematic of autoxidation leading to HOM formation.

The fate of $HOM\text{-}RO_2\bullet$ highly depends on the abundance of reaction partners. Termination reactions determine the stable HOM closed shell products.

1. $HOM\text{-}RO_2\bullet$ react with $HO_2\bullet$ produce hydroperoxide ($HOM\text{-}RO_2\bullet + HO_2\bullet \rightarrow HOM\text{-}ROOH + O_2$)
2. $HOM\text{-}RO_2\bullet$ react with other $RO_2\bullet$ lead to three different reaction channel.
    a. Alcohol and carbonyl formation ($HOM\text{-}RO_2\bullet + RO_2\bullet \rightarrow HOM\text{-}ROH$, R=O/ HOM-R=O, ROH)
    b. Alkoxy radical ($HOM\text{-}RO_2\bullet + RO_2\bullet \rightarrow HOM\text{-}RO\bullet + RO\bullet + O_2$)
    c. Accretion products ($HOM\text{-}RO_2\bullet + R'O_2\bullet \rightarrow ROOR' + O_2$)
3. $HOM\text{-}RO_2\bullet$ react with other NO lead to two different reaction channel.
    a. Alkoxy radical ($HOM\text{-}RO_2\bullet + NO \rightarrow HOM\text{-}RO\bullet + NO_2$)
    b. Organic nitrates ($HOM\text{-}RO_2\bullet + NO \rightarrow HOM\text{-}RONO_2$)

In summary, the fate of $HOM\text{-}RO_2\bullet$ peroxy radical is controlled by competition between unimolecular autoxidation and bimolecular reactions with $HO_2\bullet$, $RO_2\bullet$ and NO. Atmospheric conditions therefore strongly influence both HOM yield and product composition.

3. Detection and Identification

The observation of HOM was enabled through the development of highly sensitive chemical ionization mass spectrometry (CIMS). A widely used reagent ion is nitrate chemical ionization coupled to an atmospheric-pressure-interface time-of-flight mass spectrometer (NO_3^--CIMS or nitrate CI-APi-TOF). Nitrate ionmakes strong cluster with highly functionalized HOM which make nitrate CI-APi-TOF selective for HOM:

$HOM + NO_3^- \rightarrow HOM\cdot NO_3^-$

The resulting cluster ions can be detected by high-resolution mass spectrometry. An important analytical limitation remains: high-resolution mass spectrometry normally provides molecular formulas rather than complete molecular structures. Multiple structural isomers can have the same elemental composition but very different formation pathways and physicochemical properties.

Measurements of HOM have been performed in laboratory reactors, atmospheric simulation chambers, forests, urban environments, and remote atmospheric locations. Combining mass-spectrometric observations with chamber experiments, kinetic modeling, and quantum-chemical calculations has therefore become essential for understanding their formation mechanisms.

4. Importance for Secondary Organic Aerosol Formation

The atmospheric importance of HOM arises largely from their low volatility. Incorporation of several oxygen-containing functional groups generally reduces the vapor pressure of an organic molecule. Many HOM therefore fall within the low-volatility organic compound (LVOC) or extremely low-volatility organic compound (ELVOC) regimes.

After formation in the gas phase, sufficiently low-volatility HOM can condense onto pre-existing particles:

This process contributes directly to SOA mass. HOM can consequently explain part of the rapid SOA production observed during oxidation of monoterpenes such as α-pinene. Even lower-volatility products can contribute to the formation and early growth of molecular clusters. Organic compounds may participate together with sulfuric acid, ammonia, amines, or other atmospheric species in new particle formation. Once stable clusters have formed, condensation of HOM and other low-volatility compounds can promote their growth toward sizes at which particles may survive atmospheric coagulation losses. HOM chemistry therefore provides a mechanistic link between gas-phase VOC oxidation and atmospheric aerosol populations:

Because atmospheric aerosol particles scatter and absorb radiation and can act as cloud condensation nuclei, changes in HOM formation may ultimately influence atmospheric radiation and cloud properties.

5. Atmospheric implication

Despite substantial progress, several important uncertainties remain. First, the detailed molecular structures of many detected HOM are still unknown. Molecular formulas alone cannot identify the positions of hydroperoxide, carbonyl, hydroxyl, nitrate, or peroxide functionalities.

Second, HOM yields depend strongly on atmospheric chemical conditions. In particular, NO_x can alter RO_2 lifetimes and compete with autoxidation pathways. The balance between $RO_2 + NO$, $RO_2 + HO_2$, $RO_2 + RO_2$, and unimolecular RO_2 reactions therefore determines the distribution of oxidation products.

Third, HOM chemistry is not restricted to α-pinene. Other monoterpenes, sesquiterpenes, isoprene, and anthropogenic VOCs can also generate strongly oxygenated products, although their mechanisms and yields differ considerably.

Further integration of laboratory chamber experiments, field measurements, high-resolution mass spectrometry, theoretical calculations, and atmospheric models will be required to determine how HOM chemistry varies across atmospheric environments. Particular attention should be given to the competition between autoxidation and bimolecular RO_2 reactions because this competition controls whether carbon remains in relatively volatile gas-phase products or is rapidly converted into low-volatility material capable of forming aerosol.

6. Conclusions

Highly oxygenated organic molecules represent an important link between VOC oxidation chemistry and atmospheric aerosol formation. They are formed primarily through autoxidation sequences involving RO_2 radicals, intramolecular hydrogen shifts, and repeated O_2 addition. This chemistry enables rapid incorporation of multiple oxygen atoms into VOC oxidation products and can generate compounds with extremely low vapor pressures.

Biogenic VOCs, particularly monoterpenes such as α-pinene, are major precursors of atmospheric HOM. The subsequent competition between RO_2 autoxidation and reactions with NO, HO_2, and other RO_2 radicals determines HOM composition and yield. Once formed, low-volatility HOM can condense onto existing particles or participate in new particle formation and early particle growth.

The discovery of HOM has consequently changed the traditional picture of atmospheric VOC oxidation by demonstrating that highly oxygenated, low-volatility compounds can form within seconds rather than exclusively through multiple generations of oxidation. A quantitative understanding of this chemistry is therefore important for connecting VOC emissions with SOA formation and, ultimately, aerosol–climate interactions.

References

1. Bianchi, F., Kurtén, T., Riva, M., Mohr, C., Rissanen, M. P., Roldin, P., Berndt, T., Crounse, J. D., Wennberg, P. O., Mentel, T. F., Wildt, J., Junninen, H., Jokinen, T., Kulmala, M., Worsnop, D. R., Thornton, J. A., Donahue, N., Kjaergaard, H. G., & Ehn, M. (2019). Highly oxygenated organic molecules (HOM) from gas-phase autoxidation involving peroxy radicals: A key contributor to atmospheric aerosol. *Chemical Reviews, 119*, 3472–3509.
2. Ehn, M., Thornton, J. A., Kleist, E., Sipilä, M., Junninen, H., Pullinen, I., Springer, M., Rubach, F., Tillmann, R., Lee, B., Lopez-Hilfiker, F., Andres, S., Acir, I.-H., Rissanen, M., Jokinen, T., Schobesberger, S., Kangasluoma, J., Kontkanen, J., Nieminen, T., Kurtén, T., Nielsen, L. B., Jørgensen, S., Kjaergaard, H. G., Canagaratna, M., Dal Maso, M., Berndt, T., Petäjä, T., Wahner, A., Kerminen, V.-M., Kulmala, M., Worsnop, D. R., Wildt, J., & Mentel, T. F. (2014). A large source of low-volatility secondary organic aerosol. *Nature, 506*, 476–479.
3. Crounse, J. D., Nielsen, L. B., Jørgensen, S., Kjaergaard, H. G., & Wennberg, P. O. (2013). Autoxidation of organic compounds in the atmosphere. *The Journal of Physical Chemistry Letters, 4*, 3513–3520.
4. Jokinen, T., Sipilä, M., Richters, S., Kerminen, V.-M., Paasonen, P., Stratmann, F., Worsnop, D. R., Kulmala, M., Ehn, M., Herrmann, H., & Berndt, T. (2014). Rapid autoxidation forms highly oxidized RO_2 radicals in the atmosphere. *Angewandte Chemie International Edition, 53*, 14596–14600.
5. Rissanen, M. P., Kurtén, T., Sipilä, M., Thornton, J. A., Kangasluoma, J., Sarnela, N., Junninen, H., Jørgensen, S., Schallhart, S., Kajos, M. K., Taipale, R., Springer, M., Mentel, T. F., Ruuskanen, T., Petäjä, T., Worsnop, D. R., Kjaergaard, H. G., & Ehn, M. (2014). The formation of highly oxidized multifunctional products in the ozonolysis of cyclohexene. *Journal of the American Chemical Society, 136*, 15596–15606.
6. Berndt, T., Richters, S., Kaethner, R., Voigtländer, J., Stratmann, F., Sipilä, M., Kulmala, M., & Herrmann, H. (2015). Gas-phase ozonolysis of cycloalkenes: Formation of highly oxidized RO_2 radicals and their reactions with NO, NO_2, SO_2, and other RO_2 radicals. *The Journal of Physical Chemistry A, 119*, 10336–10348.
7. Kirkby, J., Duplissy, J., Sengupta, K., et al. (2016). Ion-induced nucleation of pure biogenic particles. *Nature, 533*, 521–526.
8. Tröstl, J., Chuang, W. K., Gordon, H., et al. (2016). The role of low-volatility organic compounds in initial particle growth in the atmosphere. *Nature, 533*, 527–531.

Mechanical Engineering Division

Welcoming Message

As a specialized division of VeKNI e.V., the Mechanical Engineering Division drives future industrial breakthroughs by integrating advanced artificial intelligence (AI) and eco-friendly paradigms into traditional mechanical and physical design. This session showcases seven cutting-edge research contributions that innovate intelligent mobility, optimize precision manufacturing, and build secure data infrastructures for a sustainable circular economy.

Focus & Core Scope

- **Aerospace & Fluid Dynamics**: Engineering biomimetic designs for low-noise propeller aerodynamics, assessing stratospheric environmental impacts (LCA) of commercial rocket launches, and developing fuel-flexible gas turbine combustion technologies using additive manufacturing.
- **Robotics & Physical AI**: Designing autonomous aerial navigation frameworks guided by 3D semantic voxel maps to safely overcome complex, high-resolution physical environments.
- **Precision Manufacturing & Optimization**: Harnessing high-precision laser processing for next-generation AI data center hardware and utilizing Bayesian optimization to minimize geometric errors in stamping hydrogen fuel cell components.
- **Digitalization & Systems Engineering**: Establishing sovereign, federated data ecosystems (Manufacturing-X) for circular economies, and proposing disciplined human-AI collaborative workflows (RPI) and verification harnesses for reliable scientific software engineering.

Learning from Nature: Biomimetic Applications of Owl-Inspired Serrations for Propeller Noise Reduction

Changjin Hong and Jong-Seob Han

Department of Aerospace Engineering, Chungnam National University, Daejeon 34134, Republic of Korea
jongseobhan@cnu.ac.kr

Abstract. Nature often solves engineering problems through distributed, multi-scale structures rather than through a single optimized component. Owl wings provide a well-known example: comb-like leading-edge serrations, a soft feather surface, and a compliant trailing edge work together to support quiet flight. This contribution introduces the biomimetic design logic behind owl-inspired serrations and connects it to a controlled propeller experiment. A 14-in, two-bladed propeller was used as the baseline, and five trailing-edge variants were produced by stereolithography. The variants differed in treated span and in whether serration height, wavelength, or both changed toward the blade tip. Aerodynamic and acoustic behavior was measured in an anechoic wind tunnel at Chungnam National University in Korea. The experiment used thrust and electrical-power measurements together with a seven-microphone arc, band-limited overall sound-pressure level, and Welch spectral analysis. The measurements indicate that biomimetic spanwise tailoring can reduce wake-side broadband noise without requiring a uniformly serrated blade. The study illustrates how observing the functional morphology of owl feathers can lead to a scalable passive-noise-control concept for drones and urban air mobility vehicles.

Keywords: Biomimicry, Owl wing, Trailing-edge serration, Propeller aeroacoustics

Introduction

Biomimetic engineering begins with a functional question: what physical problem has nature solved, and which mechanism—not merely which shape—can be transferred to technology? Owl flight is particularly instructive for aeroacoustics. Many owl species hunt using hearing and fly at low apparent noise. Their primary feathers exhibit a comb-like serrated leading edge, while the feather surface is soft and the trailing edge is fringed. These features disturb the formation and scattering of coherent turbulent structures, thereby reducing the strength and organization of aerodynamic sound sources. Importantly, the owl wing is not a uniform sawtooth plate. The morphology changes with feather position and species, suggesting that spatial distribution is part of the biological solution. Small propellers for drones and future urban air mobility vehicles face a related problem. They must generate thrust in close proximity to people while limiting tonal and broadband noise. Propeller noise contains blade-passage-frequency harmonics as well as a broadband component associated with unsteady loading and turbulent boundary-layer scattering at the trailing edge. Passive serrations are attractive because they can be integrated into the blade without active control hardware. However, a propeller rotates through strongly varying local velocity, Reynolds number, loading, and boundary-layer conditions. Applying one serration size everywhere therefore does not fully reflect the spatially varying logic observed in nature.

This article reframes the accompanying propeller study as an example of nature-inspired engineering. The objective is not to claim that an owl feather and a propeller blade are aerodynamically identical. Rather, the study transfers a design principle: tailor the edge structure to the local flow and to the region that contributes most strongly to sound radiation. The resulting question is whether spanwise-varying serrations can provide a useful acoustic benefit with less treated area and a manageable aerodynamic penalty.

Experiments and Discussion

The biological reference was documented using an open-access study of owl primary feathers. The reported morphology shows serrations concentrated toward the outer portion of the wing, with tips oriented into the incident flow. This arrangement is relevant to propellers because the outer blade has the highest tangential velocity and often makes a strong contribution to radiated noise. In the engineering translation, the owl-inspired feature was not copied at biological scale. Instead, it was represented by a controlled triangular trailing-edge serration whose dimensions could be varied systematically along the radius.

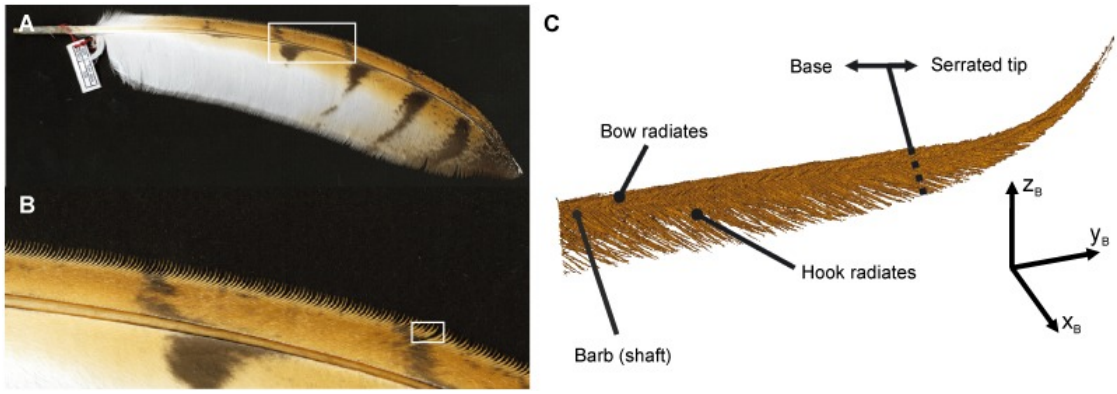


Fig. 1. Owl primary feather showing leading-edge serrations at different magnifications [1].

A two-bladed, 14-in baseline propeller was selected as the parent geometry. Its diameter was 355.6 mm, radius 177.8 mm, aspect ratio 8.5, mean chord approximately 20 mm, and hub diameter 26 mm. All test articles retained the same planform and pitch distribution; only the trailing edge was modified.

The propellers were manufactured with a Formlabs Form 4 stereolithography printer using Clear Resin V5. Each part was washed in isopropyl alcohol, post-cured, and blended at the transition between the original edge and the serrated extension. Keeping the material and post-processing identical reduced manufacturing variability. This procedure also provided a direct experimental comparison between a conventional uniform treatment and a nature-inspired spatially tailored treatment.

Experiments were conducted in a subsonic anechoic wind tunnel at Chungnam National University in Korea. The propeller was mounted on a TYTO Flight Stand 15 Pro and driven by a T-MOTOR MN3510 brushless DC motor. PWM commands were generated through an NI-9401 digital module and controlled in LabVIEW. The flight stands recorded thrust and operating data, while the electrical input system provided power-consumption data. This separation of aerodynamic and acoustic measurements allowed the serration effect to be judged as a trade-off rather than as a noise-only result.

Seven ICS-40180 MEMS microphones were arranged on a circular arc of radius 2.4 m, approximately 6.75 propeller diameters from the rotor center. The microphone polar angles were −60°, −40°, −20°, 0°, 20°, 40°, and 60°. The arc provided a directional map and enabled a compact rear-side metric based on MIC6 and MIC7, representing the wake-side sector. Before the propeller measurements, the channels were compared with a pink-noise field and adjusted against a commercial sound-level meter so that inter-channel sensitivity differences were reduced (see Fig. 2 in details).

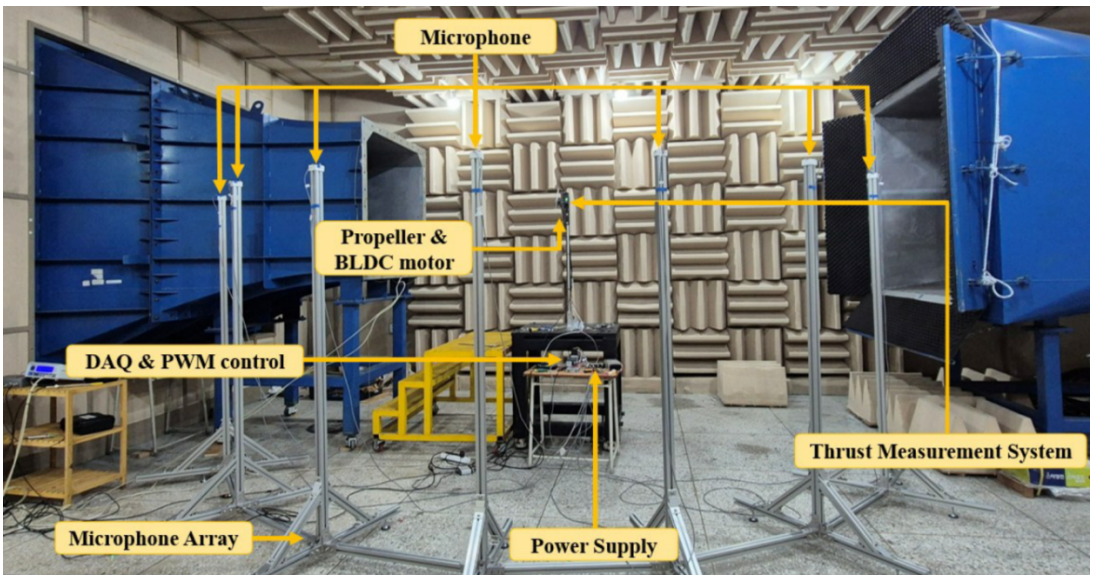


Fig. 2. Experimental setup for aerodynamic and acoustic measurements in the anechoic wind tunnel at Chungnam National University in Korea.

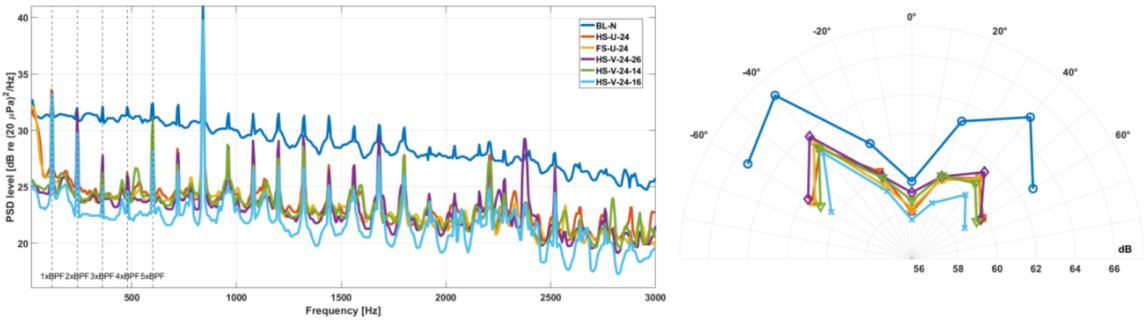


Fig. 3. Example results. Welch-averaged PSD spectra at MIC 7 and the directivity of OASPL of 6 different model propellers.

The principal outcome was a reduction of broadband noise between blade-passage-frequency harmonics, especially in the rear-side sector. Spanwise-varying cases with increasing serration height were often more favorable than the uniform half-span reference. The preferred design changed with operating condition, and the full-span uniform case remained competitive in some equal-thrust tests. Tonal changes were less consistent than broadband changes, confirming that the biomimetic serration acts most reliably as a broadband-noise-control device.

Conclusions

Owl feathers demonstrate that quiet flight can be supported by a combination of edge serration, compliant surface structure, and spatially non-uniform morphology. Translating this principle to propellers, the present study examined spanwise-varying trailing-edge serrations rather than a single uniform geometry. A controlled six-propeller matrix, identical-RPM/equal-thrust testing, a calibrated seven-microphone arc, Welch PSD processing, and thrust-normalized trade-off metrics were used to connect biological inspiration with measurable engineering performance. The short result is that appropriately tailored serrations can reduce rear-side broadband noise, sometimes with only half-span treatment. Future work should optimize the local serration dimensions together with blade loading, wake structure, and forward-flight conditions.

References

1. Weger, M., and Wagner, H.: Morphological Variations of Leading-Edge Serrations in Owls (Strigiformes). PLoS ONE 11(3), e0149236 (2016). doi:10.1371/journal.pone.0149236.
2. Candeloro, P., Ragni, D., and Pagliaroli, T.: Small-Scale Rotor Aeroacoustics for Drone Propulsion: A Review of Noise Sources and Control Strategies. Fluids 7, 279 (2022).
3. Howe, M. S.: Noise Produced by a Sawtooth Trailing Edge. J. Acoust. Soc. Am. 90, 482–487 (1991).
4. Lyu, B., Azarpeyvand, M., and Sinayoko, S.: Prediction of Noise from Serrated Trailing Edges. J. Fluid Mech. 793, 556–588 (2016).
5. Hong, C., Kim, H., Kim, J., Kim, D., and Han, J.-S.: Experimental Investigation of the Aerodynamic and Acoustic Performance of Propellers with Spanwise-Varying Trailing-Edge Serrations. Manuscript ID4464327 (2026).

Environmental Impact of Rocket Launches in New Space Era

Min Son[1]

[1]Institute of Thermodynamics, Universität der Bundeswehr München, 85577 Neubiberg, Germany
Min.Son@unibw.de

Abstract. The transition from Cold War state-led space exploration to the commercial "New Space" era has driven an unprecedented surge in orbital launch frequencies. While economic efficiency and payload miniaturization have accelerated industry growth, the associated environmental impacts remain undertreated in public policy. Simple comparisons highlighting low aggregate CO_2mass compared to commercial aviation overlook critical atmospheric dynamics. The primary environmental threats stem from the injection of highly absorptive black carbon (soot) directly into the stratosphere. Crucially, the industry's shift toward reusable rockets introduces a complex environmental paradox: retro-propulsion during re-entry requires thrust throttling, which forces engines into non-optimal mixture ratios that drastically increase soot generation in the upper atmosphere. Furthermore, we discuss European initiatives, such as Project SLICE and the European Space Agency's (ESA) Life Cycle Assessment (LCA) frameworks for launch vehicles, emphasizing the need for an eco-design paradigm shift in propulsion and system architecture.

Keywords: Space Launch Emissions, Black Carbon, Reusable Launch Vehicles, Life Cycle Assessment

1. Introduction

Following the historic space race between the United States and the Soviet Union, the global space sector is undergoing a structural paradigm shift into the "New Space" era. Driven by commercial pioneers such as SpaceX in the United States, alongside satellite miniaturization and rising demand for Low Earth Orbit (LEO) mega-constellations, space access has transitioned into a highly commercialized industry [1]. Based on industry tracking data, annual orbital launch attempts increased from approximately 60 in the early 2000s to over 160 in 2021 [2]. Venture capital investments expanded rapidly, supporting over 500 space startups by 2018 [2]. However, despite this exponential expansion, the environmental degradation caused by rocket launches has received minimal policy oversight. Public and scientific discourse began gaining substantial momentum around 2020, leading to broader media coverage, such as reports published by the BBC in July 2022 [3].

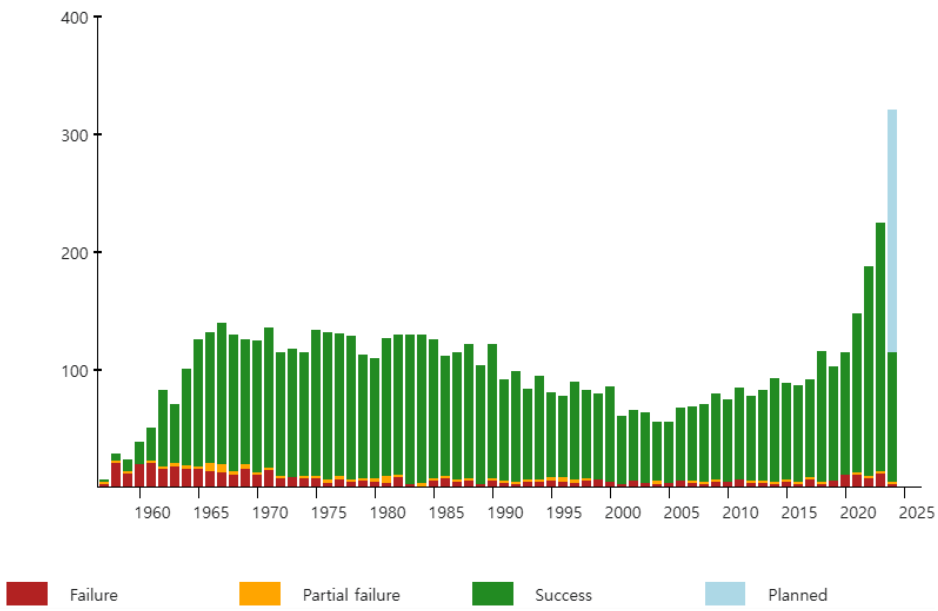


Fig. 1. Space launch statistics (1957–2026) [4]

2. Atmospheric Mechanisms

Some industry stakeholders argue that rocket emissions are negligible when compared to commercial aviation. Based on transport metrics from 2018, commercial aviation accounted for approximately 38 million flights and 900 million metric tons of CO_2, whereas rocket launches accounted for 160 flights and roughly 20,000 metric tons of CO_2 [2]. However, assessing environmental impact solely through total carbon dioxide mass is fundamentally flawed due to distinct physical and thermodynamic injection factors.

2.1 Soot (Black Carbon) and Radiative Forcing

Generally known as Black Carbon (BC) or soot, the incomplete combustion of hydrocarbon fuels (such as RP-1/kerosene) releases highly absorptive particles into the upper atmosphere. Based on research by Barker et al. published in the journal Earth's Future in 2026, rockets launching satellite mega-constellations now burn more fuel than all other types of rocket missions combined [5]. Furthermore, this rocket soot absorbs sunlight with a warming effect more than 500 times greater per unit of mass than soot released from ground-level sources like cars and power plants, and even 105 times more compared to subsonic aviation [6].

2.2 Retro-Propulsion and the Reusability Paradox

Commercial aircraft fly within the troposphere (< 12km altitude), where atmospheric circulation and rainout cleanse pollutants relatively quickly. In contrast, rockets traverse the stratosphere (10-50km) and mesosphere (50-85 km) vertically. Because the stratosphere lacks rainout mechanisms, injected species exhibit residence times ranging from months to several years [6].

Crucially, the New Space era's emphasis on Reusable Launch Vehicles (RLVs) introduces a profound environmental paradox. To recover the first stage, reusable rockets must perform retro-propulsion burns to decelerate during atmospheric re-entry. To maintain the low thrust required for a controlled descent, the engine must be heavily throttled. Based on fundamental combustion thermodynamics, operating an engine at this throttled state inherently forces the combustion into non-optimal mixture ratios (off-nominal conditions).

Generally known as a primary driver of incomplete combustion, this off-design mixture ratio drastically exacerbates the formation of soot under highly adverse conditions. If a rocket is entirely expendable, this additional re-entry burn is entirely unnecessary. Therefore, a critical question arises: Is the reusability of launch vehicles truly beneficial for the environment? While reusability saves upstream manufacturing resources, it directly penalizes the upper atmosphere by injecting disproportionately high volumes of absorptive soot during the retro-burn phase [7].

3. Regulatory Frameworks and Life Cycle Assessment

3.1 Policy Disparity

Aviation has well-established climate compliance frameworks. Based on international policy milestones, the International Civil Aviation Organization (ICAO) adopted the CORSIA framework in 2016, and the United States Environmental Protection Agency (EPA) finalized greenhouse gas regulations for aircraft in 2021[8]. In stark contrast, the space transportation sector operates without legally binding global environmental emission standards for the upper atmosphere, leaving propulsion development focused primarily on cost and payload capacity.

3.2 European Initiatives

To address these critical data gaps, European institutions are advancing rigorous research into launcher environmental impacts. Based on EU research frameworks, the initiative known as Project SLICE (Space Launch Impact on Climate and Environment) investigates the climate impacts of space launch systems by bridging space engineering and climate science. Initiated around 2026 in Europe, Project SLICE focuses on three core areas: Launch Vehicle Emissions, Atmospheric Interaction & Climate Impact, and System Analysis & Design [9].

Concurrently, under the Clean Space Initiative, the European Space Agency (ESA in Europe) is standardizing Life Cycle Assessment (LCA) methodologies specifically for space transportation [10]. Space LCA expands beyond direct launch emissions to evaluate comprehensive cradle-to-grave impacts.

- Upstream: Raw material extraction, and the highly energy-intensive manufacturing of propellants (e.g., liquid hydrogen, cryogenic processing).
- Mainstream: Launch pad construction, static fire testing, and direct ascent plume emissions.
- Downstream: Stage disposal, orbital debris generation, and atmospheric re-entry burn-up effects.

4. Conclusions

The exponential expansion of commercial spaceflight demands an urgent paradigm shift in how environmental impacts are evaluated. The argument that space launches contribute a negligible fraction of global CO_2 ignores the severe implications of black carbon in the stratosphere. Furthermore, the industry's default assumption that reusability equates to eco-friendliness must be rigorously challenged. The severe soot generation caused by non-optimal mixture ratios during re-entry thrust throttling reveals that reducing upstream manufacturing impacts may come at an unacceptable cost to the upper atmosphere [5, 7]. Environmental considerations must be integrated directly into the initial propulsion and vehicle design phase—an approach generally known as Eco-Design. Key architectural strategies must include developing green propellant combinations optimized for minimal soot across a wide range of throttle settings, and embedding comprehensive LCA frameworks early in the systems engineering process [9, 10]. As international regulatory bodies inevitably introduce space emission policies, verifiable eco-friendliness will evolve from an operational burden into a primary competitive advantage for launch service providers in the New Space era.

References

1. Ross, M., Vedda, J.A.: The policy and science of rocket emissions. Center for Space Policy and Strategy, The Aerospace Corporation 2012, 7 (2018).
2. Dallas, J.A., Raval, S., Alvarez Gaitan, J.P., Saydam, S., Dempster, A.G.: The environmental impact of emissions from space launches: A comprehensive review. Journal of Cleaner Production 255, 120209 (2020).
3. BBC News: The pollution caused by rocket launches. https://www.bbc.com/future/article/20220713-how-to-make-rocket-launches-less-polluting, last accessed 2026/08/11.
4. Timeline of spaceflight, https://en.wikipedia.org/wiki/Timeline_of_spaceflight, last accessed 2026/08/11.
5. Barker, C.R., Marais, E.A., Tan, E.Y.P., Eastham, S.D., Diskin, G.S., DiGangi, J.P., Choi, Y., Rollins, A.W., Waxman, E., Bui, T.P., et al.: Radiative forcing and ozone depletion of a decade of satellite megaconstellation missions. Earth's Future 14(5), e2025EF007229 (2026).
6. Ross, M., Mills, M., Toohey, D.: Potential climate impact of black carbon emitted by rockets. Geophysical Research Letters 34(22), (2010).
7. Larson, E.J.L., Portmann, R.W., Rosenlof, K.H., Fahey, D.W., Daniel, J.S., Langford, S.R.: Global atmospheric response to emissions from a proposed reusable space launch system. Earth's Future 5(1), 37–48 (2017).
8. Liu, Z.: Scaling sustainable propellants to address rocket emissions: Lessons from the US aviation industry. Duke Law & Technology Review 26, 28 (2026).
9. Project SLICE: Space Launch Impact on Climate and Environment. European Union CORDIS Research Programme. https://cordis.europa.eu/project/id/101227592, last accessed 2026/08/11.
10. ESA LCA Working Group, et al.: Space system life cycle assessment (LCA) guidelines. ESA Handbook ESSB-HB-U-005 (2016).

Engineering Additively Manufactured Injectors towards Fuel-Flexible and Scalable Combustion Technologies

Yeonse Kang[1*] and Fabian Hampp[1]

[1] Institute of Combustion Technology for Aerospace Engineering (IVLR), University of Stuttgart, Stuttgart, Germany
*yeonse.kang@ivlr.uni-stuttgart.de

Abstract. Additive manufacturing has evolved from a rapid-prototyping technique into an enabling technology for functional combustion components, providing new degrees of freedom in the design of fueldelivery and aerodynamic systems. This work presents an additively manufactured (AM) design-diagnostics-optimisation framework for the fuel-flexibleand scalable combustion based on a modular pilot jet-stabilised burner. Interchangeable AM components are used to modify fuel-injection concepts, nozzle and airblast geometries, swirl generation and turbulence characteristics within a jet combustion platform. Their effects are characterised using complementary optical diagnostics and quantitative data analysis, linking geometrical modifications to fuel delivery, atomization, droplet transport, mixing and flame behaviour. Previous investigations within this framework demonstrate that micro-scale tuned fuel and jet delivery concepts can provide effective atomization at comparatively low fuel-supply pressures, while aerodynamic modifications offer additional control over droplet distribution, mixing, flame stabilisation and emission level. The combined approach highlights Additive manufacturing not simply as a manufacturing route, but as a means of systematically exploring coupled injector-flow design spaces. Such integration of manufacturing, diagnostics and performance optimisation provides a pathway towards fuel-flexible and scalable gas turbine combustion technologies.

Keywords: Additive manufacturing; Fuel injection; Spray atomization; Laser diagnostics; Fuel flexibility; Gas turbine combustion

1. Introduction

The idea of creating structures by progressively adding material is far older than what is now termed additive manufacturing. From masonry structures assembled layer by layer to modern engineering components, building by addition represents a simple principle: material is placed where it is needed to create the desired form. Modern additive manufacturing (AM) transformed this principle by combining layer-wise fabrication with digital design and precise control of material deposition. What once operated predominantly at the scale of structures can now be applied to metallic components containing intricate internal features at the sub-millimetre scale [1, 2].

This transition has particular significance for fuel injectors. Conventional manufacturing can produce injector orifices on the order of 100 μm; for example, spray holes with diameters of 100–200 μmand depths of 0.7–1.1mm have been manufactured using established machining routes [3, 4]. However, further miniaturisation and increasing geometrical complexity impose substantial constraints on machining access, tool dimensions, aspect ratio and manufacturing cost. Metal AM, and particularly laser powder bed fusion (LPBF), relaxes some of these constraints while enabling part consolidation and more efficient material utilisation. More importantly for fuel injection, it allows complex internal flow passages and fuel outlets to become active design variables rather than manufacturing constraints. This expanded geometrical freedom provides new means of controlling fuel delivery and atomization, with subsequent implications for the scalability and fuel flexibility of combustion nozzle designs.

The potential of this design freedom is already evident in modern gas-turbine (GT) technology. Additive manufacturing has progressed from the fabrication of prototypes to the production and repair of functional components, including fuel nozzles, burners and components incorporating complex cooling passages [4-7]. A prominent example is the AM fuel nozzle developed for the LEAP aero-engine, where AM enabled the consolidation of a conventionally assembled multi-part design into a single component [8]. In stationary gas turbines, AM has similarly been utilised to manufacture and redesign burner components, exploiting complex internal geometries and integrated functionalities that would be difficult to realise through conventional manufacturing [9, 10]. These applications demonstrate the technological maturity of AM for severe GTenvironments, while also highlighting its potential to reconsider the design of the fuel-delivery system itself.

Building on these capabilities, our research explores how AM can be used not only to manufacture combustion components, but also to systematically expand and exploit the design space of fuel injection and flame stabilisation systems. As illustrated in Fig. 1, a modular pilot jet-stabilised burner provides a common platform in which AM enables rapid variation of nozzle structures, fuel-injection concepts, swirl generation and turbulence-control elements. Their influence on fuel delivery, atomization, mixing and flame behaviour is subsequently characterised using complementary laser and optical diagnostics, together with quantitative data-processing methods. This integrated design framework provides a basis for identifying the geometrical and operating parameters governing combustion performance and for progressively optimising the injector towards enhanced atomization, stable and clean combustion, and reduced pollutant emissions

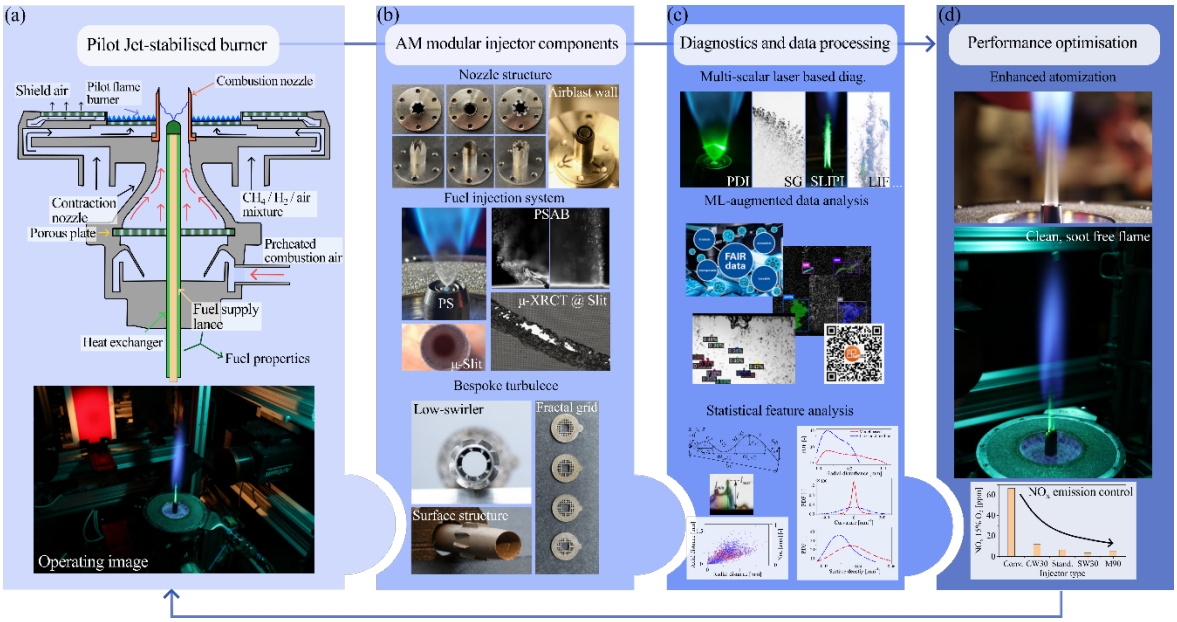


Fig. 1. AM-enabled design, characterisation and optimisation framework for the pilot jet-stabilised combustion system. (a) Pilot jet-stabilised burner architecture, showing the central fuel-supply lance, pilot burner, contraction nozzle, porous plate, heat exchanger and combustion nozzle, together with a representative operating flame. (b) AM-enabled modular design space, including interchangeable nozzle head structures, airblast-wall geometries, conventional and AM fuel-injection concepts, low-swirl elements, fractal grids and μ-scale slit injectors. (c) Experimental characterisation and data-processing framework combining multi-scalar laser-based diagnostics, including PDI, SG/SLIPI and LIF, with multi-augmented and statistical feature analyses. (d) Translation of the resulting design and diagnostic knowledge towards performance optimisation, targeting enhanced atomization, clean and soot-free flame operation, and reduced NO_X emissions.

2. Experiments and Discussion

The experimental investigations are conducted using the modular pilot jet-stabilised burner shown schematically in Fig. 1(a), together with a representative image of the burner under reacting operation. The facility provides independent control of the pilot flame, preheated combustion air and liquid and gaseous fuel supplies, allowing different fuelinjection and flamestabilisation concepts to be investigated within a common combustion platform. Representative liquid-fuel investigations use Jet A-1 with fuel mass flow rates of 0.1– 5 g/s, combustion-air jet velocities of 60 –160 m/s and preheat temperatures of 300– 800 K, while the modular fuel-delivery system also permits operation with gaseous fuels, such as hydrogen, methane, and bio-gas.As illustrated in Fig. 1(b), AM is utilisedto realise interchangeable injector and flow-conditioning components, including different nozzle structures, airblast geometries, μ-scale slit fuel-injection concepts, swirlers and turbulence-generating elements. Their effects on fuel delivery, spray formation and combustion arecharacterised using complementary optical diagnostics, including Phase Doppler Interferometry (PDI), Shadowgraphy (SG), Structured Laser Illumination Planar Imaging (SLIPI), laser-induced fluorescence (LIF), flame proxy chemiluminescence imaging, and direct RGB imaging.

The modularity of the burner is exploited to investigate different AM-enabled strategies for controlling fuel delivery and the surrounding aerodynamic field. At the injector level, interchangeable components allow the nozzle structure and airblast geometry to be varied, while fuel-injection concepts ranging from conventional pressure-swirl and airblast injectors to AM μ-slit configurations are investigated [11–15]. Beyond the injector itself, AM enables aerodynamic modifications at increasingly small geometrical scales, including low-swirl elements with minimum vane thicknesses of approximately 80 μm, fractal grids and tailored surface structures with characteristic feature heights of approximately 40 μm [11–13]. The low-swirl configurationsincrease the near-injector turbulent kinetic energy by up to approximately 135% and improve the stability of radial fuel placement by approximately 50%, demonstrating how geometrical modifications can substantially reorganise the coupled gas-liquid flow [11,12]. Meanwhile, small-scale turbulence modulation using grids and tailored surface structures perturbs the shear layer, thereby modifying the flame-onset characteristics [14]. At a still smaller scale, LPBF enables annular μ-slit fuel injectors with a measured gap width of 52 μm with XRCT [14,15]. The resulting annular fuel delivery produces droplets with the size on the order of 10 μm while maintaining a fuel-pressure drop below 8 bar, providing a wide operating range and the fast dynamic response with comparatively low hydraulic loss [14,15]. The more uniform circumferential fuel placement further promotes interaction between the liquid phase and the surrounding air jet streams, with subsequent benefits for flame stabilisation and pollutant emissions [14,15]. Collectively, these examples demonstrate that AM provides not only a route for manufacturing individual components, but also a means of systematically varying the geometrical, aerodynamic and fuel-delivery boundary conditions governing spray formation and transport, while providing the flexibility required to accommodate different fuels and combustion-system scales.

The resulting design space requires complementary diagnostics capable of resolving how geometrical modifications propagate through the spray and combustion process. As summarised in Fig. 1(c), Phase Doppler Interferometry (PDI) provides quantitative droplet size and velocity statistics, while shadowgraphy (SG) and Structured Laser Illumination Planar Imaging (SLIPI) resolve the spatial organisation of the liquid phase. Laser-induced fluorescence (LIF) and chemiluminescence imaging further provide information on fuel transport and the resulting flame response [15,16]. These measurements are complemented by quantitative image processing, statistical feature extraction and machine-learning (ML)-based object detection and segmentation, enabling relevant spray and flame structures to be identified and compared systematically across the AM design space. Rather than treating individual measurements independently, the combined diagnostic framework traces the response from the imposed geometry and aerodynamic boundary conditions through atomization, droplet transport and fuel–air mixing to flame stabilisation and pollutant formation. As illustrated in Fig. 1(d), the resulting spray and combustion metrics are then evaluated against performance objectives, including atomization quality, flame stability and pollutant emissions [11,15]. This closes the design loop by feeding the experimentally identified relationships back into the selection and optimisation of injector geometry, fuel-delivery concepts and aerodynamic control elements. The framework therefore progresses iteratively from AM-enabled design through quantitative diagnostics and data analysis to combustion-performance assessment and subsequent design optimisation. Such an approach is particularly relevant to gas-turbine combustion, where compact injector architectures must accommodate changing fuel properties and operating conditions while simultaneously satisfying requirements for stable operation, low emissions and scalability across different power classes. The integration of AM-enabled geometrical freedom with experimentally informed optimisation therefore provides a pathway towards the development of future fuel-flexible gas-turbine combustion systems.

3. Conclusions

Additive manufacturing not only provides an alternative route for producing fuel injectors but also expands the design space through which fuel delivery, atomization and the surrounding aerodynamic field can be tailored. The modular combustion platform presented here demonstrates how this geometrical freedom can be systematically combined with quantitative optical diagnostics to establish relationships between injector design, spray formation, fuel transport and combustion performance. The investigated concepts highlight the complementary roles of micro-scale fuel delivery for effective atomization and aerodynamic modifications for controlling subsequent droplet transport and mixing, ultimately influencing flame stabilisation and pollutant formation. Future developments should therefore consider the injector, aerodynamic field and fuel properties as a coupled system, while accounting for manufacturing-induced deviations as part of the actual boundary conditions. In this context, AM becomes not simply a manufacturing technique, but an enabling element of an integrated design, diagnostics, and optimisation methodology towards fuel-flexible and scalable combustion technologies.

References

1. A. Vafadar, F. Guzzomi, A. Rassau, K. Hayward, Advances in metal additive manufacturing: A review of common processes, industrial applications, and current challenges, Applied Sciences 11 (2021) 1213.
2. M. Yakout, M.A. Elbestawi, S.C. Veldhuis, A review of metal additive manufacturing technologies, Solid State Phenomena 278 (2018) 1–14.
3. T. Pereira, J.V. Kennedy, J. Potgieter, A comparison of traditional manufacturing vs additive manufacturing, the best method for the job, Procedia Manufacturing 30 (2019) 11–18.
4. A.W. Gebisa, H.G. Lemu, Additive manufacturing for the manufacture of gas turbine engine components: Literature review and future perspectives, Proceedings of ASME Turbo Expo (2018) GT2018-76686.
5. F. Giuliani, N. Paulitsch, D. Cozzi, M. Görtler, L. Andracher, An assessment on the benefits of additive manufacturing regarding new swirler geometries for gas turbine burners, Proceedings of ASME Turbo Expo (2018) GT2018-75165.
6. R. Rajasegar, C.M. Mitsingas, E.K. Mayhew, Q. Liu, T. Lee, J. Yoo, Development and characterization of additive-manufactured mesoscale combustor array, Journal of Energy Engineering 144 (2018) 04018013.
7. A. Durocher et al., Characterization of a novel additive manufacturing micromix nozzle burning methane to hydrogen, Journal of Engineering for Gas Turbines and Power 146 (2024) 051009.
8. T. Saracyakupoglu, The qualification of the additively manufactured parts inthe aviation industry, American Journal of Aerospace Engineering6(2019) 1-10.
9. Siemens, Siemens achieves breakthrough with 3D-printed combustion component for SGT-A05, https://press.siemens.com/global/en/feature/siemens-achieves-breakthrough-3d-printed-combustion-component-sgt-a05, last accessed 2026/08/12.
10. GE Aerospace, Manufacturing Milestone: 30,000 Additive Fuel Nozzles,https://www.geaerospace.com/news/articles/manufacturing/manufacturing-milestone-30000-additive-fuel-nozzles, last accessed2026/08/12.
11. Y. Kang, J. Ahn, F. Hampp, Low swirl effect on compact spray and combustion systems using additive manufactured dual airblast injectors, Journal of Engineering for Gas Turbines and Power 146 (2024) 121001.
12. Y. Kang, F. Hampp, Low-swirl effects on spray flames for compact jet-stabilised combustion systems, Combustion and Flame 285 (2026) 114740.
13. Y. Kang, J. Seidler, J. Ahn, V. Rubio, C. Maucher, H.-C. Möhring, F. Hampp, AM micro-structures with bespoke permeability, International Journal of Heat and Mass Transfer 241 (2025) 126674.
14. Y. Kang, O. Lammel, M. Ruf, H. Steeb, H.-C. Möhring, F. Hampp, Additive manufacturing enabled annular μ-slit injection in low-NO_X jet-stabilised liquid fuel combustion, Applications in Energy and Combustion Science 27 (2026) 100519.
15. Y. Kang, B. Jose, F. Hampp, Pushing fuel flexibility with AM μ-slit injector for jet-stabilised combustion, International Journal of Spray and Combustion Dynamics 18 (2026) 26–40.
16. B. Jose, D. Greenblatt, R.P. Lindstedt, A. Breicher, D. Geyer, O. Lammel, F. Hampp, Transferable DL for in-situ validated LIF segmentation, Proceedings of the Combustion Institute 45 (2026) 106253.

Towards Learned Search Guidance for 3D Semantic Voxel Maps for Aerial Navigation

Taehyoung Kim

Fraunhofer Institute for Transportation and Infrastructure Systems IVI, Zeunerstrasse 38, 01069 Dresden, Germany
taehyoung.kim@ivi.fraunhofer.de

Abstract. Autonomous aerial navigation necessitates three-dimensional environment representations for safe path planning. Voxel maps are attractive for near-ground flight as they preserve the vertical structure of obstacles and can associate semantic information with occupancy. Yet, extending learning-based path guidance from 2D grids to such representations introduces two practical problems: the learning pipeline must operate over a much larger volumetric search space, and suitable semantic 3D training data are difficult to obtain at scale. In this work, we explore both aspects as a preliminary extension of FlexPath to 3D voxel maps from 2D grids. We demonstrate how a dense learned path prior and its objective-based adaptation can be transferred to volumetric inputs and discuss a direct 3D U-Net implementation together with its scalability limitations. We further consider two routes for generating training environments: full semantic 3D reconstruction from aerial imagery and a more scalable 2.5D approximation based on top-down semantics and height maps. A qualitative example illustrates the generation of RGB and semantic voxel representations of the same aerial scene.

Keywords: Aerial Navigation, Path Planning, Learned Guidance, Semantic Voxel Maps.

1. Introduction

Autonomous aerial systems require an environment representation that can be queried for collision checking, planning, and spatial reasoning. For near-ground flight, this representation is inherently three-dimensional: buildings, vegetation, overhangs, and passages may occupy the same horizontal location at different heights. Voxel maps provide a convenient interface to this geometry and have been used for onboard aerial planning [1]. When semantic information is fused into the same map, planning can also reason about what occupies the space rather than only whether it is free or occupied.

Classical planners such as A* remains strong for finding efficient collision-free shortest paths [2]. However, dense 3D voxel grids greatly increase the number of search states, especially at high resolution, thereby limiting the scalability of classical planners. Furthermore, in low-altitude flight, the shortest path is not always the most desirable path. Depending on the task, a UAV may need to keep a distance from obstacles, avoid no-fly zones, and remain close to areas of interest during inspection. Conventional planners encode these preferences as manually tuned costs for factors such as obstacle clearance [3].

Learning-based search guidance offers an alternative approach. A model predicts path regions in the search space, while a conventional planner still extracts the final valid path. Neural A* [4] and TransPath [5] demonstrated this idea for 2D grid-based planning and FlexPath [6] further improves it using a learned connectivity prior from task-dependent Path Shape Objectives (PSOs), allowing the prior to be adapted toward clearance, semantic preferences, or waypoint constraints.

Here, we examine what is required to transfer this idea to 3D aerial planning. We consider two coupled questions: how the 2D learning pipeline can be adapted to voxel inputs, and how sufficiently realistic semantic 3D training environments can be generated. The aim is to provide a proof-of-concept and design study rather than a full quantitative claim against established 3D planners.

2. Extending Learned Guidance to 3D

2.1 Model and Objective Adaptation

FlexPath predicts a soft spatial prior rather than the final discrete path. In 2D, the network produces an $H \times W$ confidence map that highlights cells likely to support a connection between the start and the goal [6]. As

Figure 1 depicts, a direct 3D extension produces an H x W x D confidence volume. The interpretation remains the same: learning narrows or reshapes the search region, while graph search is still responsible for finding a valid path.

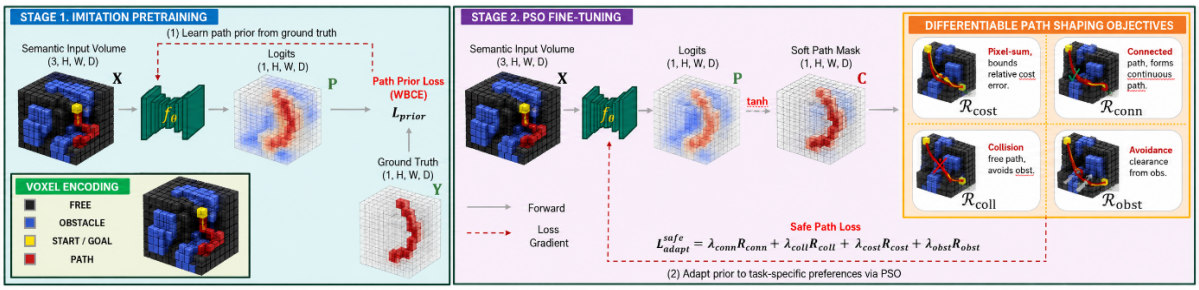


Fig. 1. Transfer of the FlexPath architecture from a 2D to 3D. A direct proof-of-concept implementation replaces the 2D U-Net and local PSO operators with their volumetric counterparts, while leveraging graph search to extract the final path.

For a first implementation, the architectural changes can be kept simple: a 2D U-Net becomes a 3D U-Net, connectivity operations use local 3D neighborhoods, and clearance is evaluated with the 3D distance to occupied voxels. Semantic objectives can operate directly on voxel labels, for example, to assign larger margins to selected classes. This direct transfer is useful for testing the idea, but it also exposes the main difficulty of the 3D setting: dense feature volumes grow with depth as well as image height and width.

2.2 Constructing Semantic 3D Training Environments

Model adaptation is only half of the problem. Learned guidance also needs many planning scenes with realistic geometry and semantics. Existing aerial 3D datasets are generally built for reconstruction, segmentation, or occupancy prediction rather than start-goal-conditioned planning; OccuFly [7], for example, provides semantic aerial occupancy data but not FlexPath-style path supervision. For training, a 3D scene must therefore be paired with sampled start-goal states and planner-generated path labels.

A high-fidelity route starts from geo-referenced nadir and oblique UAV images. Structure-from-Motion and multi-view stereo reconstruct the scene geometry, while semantic predictions from the source images are projected onto and fused into the reconstruction prior to voxelization. The same reconstructed scene can retain RGB appearance and a semantic label volume, as illustrated in Figures 2(a) and 2(b). Once free space is available, start and goal positions can be sampled automatically and 3D A* or Dijkstra search can provide connectivity or path supervision.

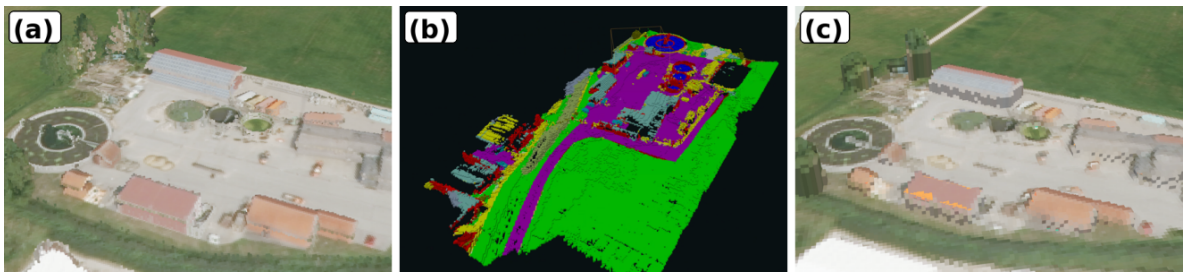


Fig. 2. Example representations used for constructing semantic 3D training environments. (a) RGB voxel reconstruction of an aerial scene. (b) Semantic voxel representation of the reconstructed scene. (c) Example of a scalable 2.5D representation generated from top-down imagery and height information.

For more extensive geographic coverage, a cheaper alternative is to combine satellite or orthophotography imagery with a digital terrain model (DTM) and digital surface model (DSM). Top-down semantic predictions can be assigned to each horizontal location and extruded over the estimated occupied height. This can generate many approximate environments, but the result is fundamentally 2.5D: all occupied voxels in one vertical column inherit the same top-down class. It cannot describe a road below a bridge, ground below a vegetation canopy, or other semantic changes along the same vertical pillar. Figure 2(c) demonstrates an exemplar case of satellite imagery and a digital terrain model. A practical compromise may therefore be to

pretrain on many such approximate scenes and adapt to a smaller set of fully reconstructed semantic 3D environments.

3. Qualitative Example and Limitations

Figures 2(a) and 2(b) illustrate the intended output of the high-fidelity mapping pipeline, in which the same aerial reconstruction is represented in two forms: one with RGB appearance and another using semantic voxel labels. This intermediate representation is important because the planning model requires both accurate 3D geometry and consistent semantic information. Unlike a top-down extrusion approach, a full volumetric reconstruction preserves vertical structures and allows different semantic classes to be assigned at different heights. Figure 3 presents a qualitative example of the planner operating on a (42 x 42 x 42) voxel grid. The model is trained with an obstacle-clearance objective.

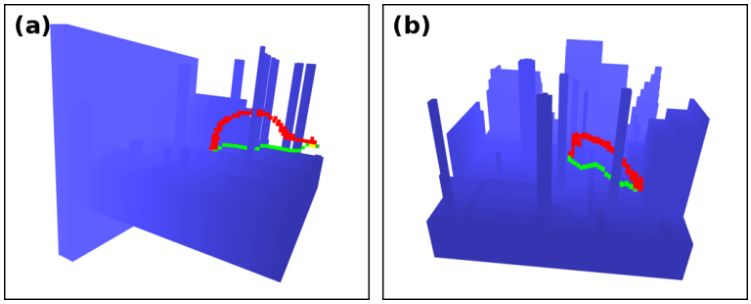


Fig. 3. Qualitative proof-of-concept planning results for a model trained with an obstacle-clearance objective. The predicted path is shown in red and the A* path in green. Blue voxels represent obstacles, and void voxels represent free space.

A major limitation of directly extending a 2D approach to 3D is the scalability. While a dense 2D network operates on (H x W) cells, a volumetric network processes (H x W x D) voxels, maintaining three-dimensional feature representations throughout the encoder and decoder. Consequently, a conventional 3D U-Net can become computationally and memory-intensive as resolution increases, and the corresponding search space grows substantially. Therefore, the dense 3D U-Net used in this study requires more efficient architectural changes and is not a scalable solution. More efficient approaches, including sparse convolutions, localized spatial tiles, and multi-resolution representations, provide promising directions for future work.

A similar trade-off exists in the data-generation pipeline. High-fidelity photogrammetric semantic reconstruction provides richer and more detailed supervision, but its acquisition and processing are relatively expensive. In contrast, top-down 2.5D representations are considerably easier to generate and scale, but they can introduce systematic errors when vertical structure plays an important role in navigation. The current study is therefore primarily qualitative. A systematic evaluation of computational runtime, node expansion, path quality, and the effectiveness of transferring knowledge from 2.5D pretraining to genuine 3D environments remains an important area for future investigation.

4. Conclusions

We explored extending learned path guidance from 2D grids to semantic 3D voxel maps from two perspectives: model adaptation and training-data generation. The FlexPath formulation transfers conceptually to volumetric inputs, but a direct dense 3D U-Net introduces a significant scalability problem. At the same time, realistic semantic 3D training environments remain expensive to generate. Full photogrammetric reconstruction and scalable top-down 2.5D map generation therefore provide complementary data sources. Future work will focus on sparse or multi-resolution architecture and on evaluating whether priors learned from approximate 2.5D scenes transfer to fully reconstructed 3D environments.

References

[1] Oleynikova, H., Taylor, Z., Fehr, M., Nieto, J., Siegwart, R.: Voxblox: Incremental 3D Euclidean Signed Distance Fields for On-Board MAV Planning. IEEE/RSJ IROS (2017).

[2] Hart, P. E., Nilsson, N. J., Raphael, B.: A Formal Basis for the Heuristic Determination of Minimum Cost Paths. IEEE Trans. Systems Science and Cybernetics 4(2), 100-107 (1968).

[3] Triest, S., Castro, M.G., Maheshwari, P., Sivaprakasam, M., Wang, W., Scherer, S.: Learning Risk-Aware Costmaps via Inverse Reinforcement Learning for Off-Road Navigation. IEEE ICRA, 924–930 (2023).

[4] Yonetani, R., Taniai, T., Barekatain, M., Nishimura, M., Kanezaki, A.: Path Planning using Neural A* Search. ICML, PMLR 139, 12029-12039 (2021).

[5] Kirilenko, D., Andreychuk, A., Panov, A., Yakovlev, K.: TransPath: Learning Heuristics for Grid-Based Pathfinding via Transformers. AAAI 37(10), 12436-12443 (2023).

[6] Kim, T., Schoenbrod, T., Eckel, D., Meess, H.: FlexPath: Learned Semantic Path Priors for Image-Based Planning. arXiv:2606.10167 (2026).

[7] Gross, M., Matha, S. B., Fahmy, A., Song, R., Cremers, D., Meess, H.: OccuFly: A 3D Vision Benchmark for Semantic Scene Completion from the Aerial Perspective. CVPR (2026)

The Hidden Technology Behind Data Centers: Laser Processing

Woo-Sik Chung[1]

[1]TRUMPF Laser-& Systemtechnik SE, Johann-Maus-Str. 2, 72541Ditzingen, Germany

woo-sik.chung@trumpf.com

Abstract. Global investment in AI driven data centers is accelerating. The physical infrastructure behind them faces mounting demands for speed, precision, and scale. High speed interconnects now transmit up to 200 Gbit/s. This places tight tolerances on the millions of copper contacts and connectors that form a data center's internal highways. This article introduces three laser-based manufacturing processes that address these challenges. Ultrashort pulse lasers selectively remove cable insulation layers with contact free precision. They replace multi station mechanical methods and enable leaner production lines. Green wavelength lasers combined with AI based vision inspection achieve stable, high yield welding of copper conductors to connectors. This combination increases productivity while keeping yields above 99 percent. Laser processing also supports the production of cold plates for liquid cooling systems. Together, these applications show how laser technology underpins the reliability and scalability of next generation data center infrastructure.

Keywords: Laser material processing, data center, vision system

1. Introduction

Artificial intelligence is reshaping the world at remarkable speed. Behind every AI model, every cloud application, and every real time data analysis lies a physical infrastructure that rarely gets noticed. That infrastructure is the data center. Global players like AWS, Google, and Meta are pouring enormous sums into new capacity. Worldwide investment in data centers is expected to exceed €670 billion by 2030.As data centers grow, so do the demands on internal communication. High-performance servers, GPU clusters, and storage systems must exchange massive volumes of data with virtually no delay. The latest high speed interconnect standards transmit up to 200 gigabits per second. That is roughly 2,000 times faster than a typical home internet connection.

Keeping this data flowing reliably depends on getting every small detail right. Millions of copper contacts, connectors, and cable joints form the physical highways of a modern data center. Even the smallest manufacturing deviation can degrade signal quality. At the same time, components keep shrinking to pack more transmission capacity into less space. Connector manufacturers face a dual challenge. They must achieve the highest quality while scaling production economically. This is exactly where modern laser technology comes in. This article looks at three laser processes currently used in the manufacturing of data center components.

2. Cable Insulation Removal — Ultrashort-Pulse Lasers

Before high-speed cables can be reliably joined, they must be precisely prepared. Modern data cables consist of multiple conductors, insulation layers, and functional layers. These layers protect the cable, preserve electrical performance, and prevent interference between individual wires. At the point where the connection will actually be made, these layers must be selectively removed. This exposes the bare copper conductor underneath. As data rates climb and components continue to shrink, this step becomes increasingly demanding. Many individual wires, so called differential pairs, run tightly packed next to one another. The material must be removed with pinpoint accuracy, exactly where the weld will happen. It cannot damage the copper conductor or the surrounding insulation. Manufacturing has traditionally relied on two approaches:

- **Mechanical removal** cuts, peels, or strips the material away. Mechanical tools act directly on the part. They can leave burrs, loose fibers, or ragged cut edges. Soft metals like copper are particularly at risk

of being damaged in the process. Plastics do not always separate cleanly this way either. For high speed connections this is a serious problem. Any damage, deformation, or contamination can affect contact quality and ultimately signal transmission.

- **Laser ablation** has historically required more complexity than necessary. Removing several different materials, insulation and conductor alike, often required multiple laser sources and several separate processing stations. That added complexity to the production line. It made it difficult to manufacture high volumes economically and reliably.

A newer, leaner approach has changed that. Ultrashort pulse lasers can remove insulation layers with high precision. The process is contact free and selective. Layers come off one at a time without affecting the surrounding material. The short pulses deposit energy at an extremely targeted point. Crucially, a single laser source can now handle all the relevant material layers. The benefits for manufacturing are substantial. The equipment footprint shrinks, since fewer stations mean fewer hand offs, fewer alignment steps, and simpler automation. Process stability improves, since the ablation is reproducible and contact free with high spatial resolution. Quality at the prepared contact point improves too, since the copper conductor is exposed without mechanical stress. Laser based ablation has effectively become an enabler for the series production of modern high-speed connectors. It sets the stage for the next step: precisely welding the copper wire to the contact surface.

3. Connector Welding — Green Lasers and AI-Based Vision Inspection

Once the conductors and contact surfaces are prepared, the copper conductor must be joined to the connector. For data to flow reliably, these electrical connections have to meet extremely tight tolerances. Even the smallest irregularity or positional deviation can degrade signal quality and lead to data loss. The challenge is that copper is notoriously difficult to weld. It reflects a large portion of laser energy. This makes conventional infrared laser processes prone to instability. To address this, the industry has turned to lasers in the green wavelength range. Copper absorbs green light far more efficiently than infrared. The result is a more stable weld pool and a more reproducible process overall. In practice, this approach has been shown to boost productivity by up to tenfold compared with conventional methods. Yields remain above 99 percent. The narrower heat affected zone also protects nearby components, such as plastic housings, from unwanted thermal stress.

As components continue to miniaturize, accurately detecting them matters more than ever. So does automatically compensating for process deviations. This is where AI based vision inspection comes in, working alongside the laser welding process. Before welding, cameras capture the components and analyze their position. Specially trained AI models identify the relevant geometry. They detect gaps or misalignments and determine the optimal weld position. The system then automatically adjusts the process parameters. This ensures the laser hits exactly the right spot even when there are minor part tolerances. This vision system can detect components and gaps within milliseconds. Quality control happens directly within the running production process. After welding, the system reanalyzes the resulting joint. It evaluates the connection and classifies it as good or defective. Quality issues are caught immediately. Process data remains available for traceability and continuous improvement.

4. Cold Plate Manufacturing — Precision Laser Machining

As computing power increases, so does the amount of heat that needs to be removed. Air cooling has long been the standard approach for servers, but it is now approaching its physical limits. As a result, data center operators are increasingly turning to liquid cooling systems built around so called cold plates. These metal heat sinks dissipate heat far more efficiently. Coolant flows through finely structured internal channels. The challenge is that these cooling channels must meet increasingly complex geometric requirements. At the same time, they must deliver high leak tightness, material quality, and thermal performance.

This is where laser-based manufacturing technologies open up new possibilities. Lasers allow for highly precise machining of metal components. They can produce complex geometries with excellent repeatability. They also offer potential for entirely new manufacturing approaches. These approaches could make cooling structures more efficient to produce and further boost cold plate performance. This is an area of laser technology development that is actively progressing today.

5. Conclusions

The global expansion of AI infrastructure is presenting the industry with new challenges across the board. These span semiconductor manufacturing, interconnect technology, cooling, and the structural and electrical infrastructure of modern data centers. What matters most is not the performance of any single component. It is how effectively the entire value chain works together.

The three examples covered here are insulation removal, copper connector welding, and cold plate machining. All three illustrate how laser technology is being applied at small but critical points throughout the vast infrastructure of a data center. Ultimately, the future of AI will not be defined by more powerful chips alone. It will depend on how manufacturing technology, interconnect technology, automation, and infrastructure come together. Laser processing is one of the quieter links in that chain. It is easy to overlook, but essential all the same.

Design of Stamping Tool Geometry for Metallic Bipolar Plates Using Bayesian Optimization

LANGENSTÜCK-LEE Sangwook[1], PARK Hyunjin[1], HÜSING Florian[1] and BRECHER Christian[2]

[1]Fraunhofer Institute for Production Technology IPT, Steinbachstraße 17, 52074 Aachen, Germany

[2]Laboratory for Machine Tools and Production Engineering (WZL) of RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany
sangwook.langenstueck-lee@ipt.fraunhofer.de

Abstract. One of the technical challenges in the manufacturing of bipolar plates (BPPs), a key component of Proton Exchange Membrane Fuel Cells (PEMFCs), is minimizing geometric deviations caused by springback after forming. These deviations can affect the dimensional accuracy of BPPs and influence the performance of the fuel cell stack. This study aims to improve the forming accuracy of metallic BPPs by optimizing the forming tool geometry using a Bayesian optimization framework. Finite element (FE) simulations were performed in ABAQUS to evaluate the resulting BPP geometry for different tool parameters, and Bayesian optimization was applied to efficiently identify optimal tool geometries while reducing the required number of simulations. Compared with a conventional Design of Experiments (DoE) approach, the Bayesian optimization-based optimization achieved a lower shape error while reducing the number of simulation iterations by 76.74%. Furthermore, an automated workflow was developed by integrating FE simulation, geometric evaluation, and optimization processes.

Keywords: Bipolar plates (BPPs), Stamping, Tool compensation, Bayesian optimization, Design of experiment, Proton Exchange Membrane Fuel Cells (PEMFCs), Springback

1. Introduction

The growing demand for environmentally friendly energy and increasing concerns over climate change have highlighted the importance of developing technologies for decarbonization [1]. In this context, Proton Exchange Membrane Fuel Cells (PEMFCs) have emerged as a promising solution due to their high energy efficiency, high power density, and potential for automotive and portable applications [2].

Bipolar plates (BPPs), one of the key components of PEMFCs, play essential roles in distributing reactant gases, conducting electrical current, managing thermal and moisture conditions, and providing structural support for the fuel cell stack [3]. The performance of PEMFCs is strongly influenced by the geometric design and manufacturing accuracy of BPPs. Therefore, designing appropriate channel geometries and ensuring their precise manufacturing remain critical challenges [4]. In recent years, micro-scale manufacturing technologies for BPPs with increasingly complex and fine structural features have attracted considerable attention as a means of achieving higher power density while reducing the weight of metallic BPPs [5, 6].

Among various manufacturing processes, stamping is widely used for metallic BPPs due to its high productivity and cost-effectiveness in mass production [7]. However, geometric deviations caused by springback of thin metal sheets are unavoidable and remain a significant challenge [8]. Such deviations can directly affect the uniformity of contact pressure, thereby degrading the performance and durability of PEMFCs and increasing the risk of gas leakage [9]. Therefore, compensation of the forming tool geometry to account for springback is essential for ensuring the geometric accuracy of the final BPP.

As a baseline approach, Design of Experiments (DoE) was employed to identify suitable tool geometry parameters. Although DoE provides a systematic approach for investigating the relationships between multiple input variables and corresponding output responses, it generally requires a large number of simulations when exploring the design space, resulting in considerable computational cost. To overcome this limitation, this study proposes a Bayesian optimization-based framework for tool geometry compensation. Bayesian optimization is particularly suitable for optimization problems involving computationally expensive

and black-box objective functions. By constructing a surrogate model from previously evaluated data and using an acquisition function to determine promising candidates for subsequent evaluations, Bayesian optimization enables efficient exploration of the design space with a reduced number of simulations.

Furthermore, an automated workflow was developed by integrating finite element (FE) simulation in ABAQUS, geometric evaluation, and optimization into a unified framework. The objective of this study is to improve the manufacturing accuracy of micro-scale metallic BPPs for PEMFCs by efficiently optimizing the forming tool geometry using Bayesian optimization.

2. Experiments and Discussion

FE simulations were conducted using ABAQUS/CAE 2023 to investigate the forming and springback behavior of metallic BPPs. The punch and die were modeled as rigid bodies, and a 0.1-mm-thick EN 1.4404 stainless steel sheet was used as the workpiece. A forming pressure of 135 MPa was applied during the stamping process. The cross-sectional geometry of the BPP channel was characterized by channel height(H), top edge length(L1), bottom edge length(L2), top fillet radius(R1), bottom fillet radius(R2) and slope angle(α) as described in Figure 1. The pitch(p) was fixed to maintain the required geometric compatibility of the BPP.

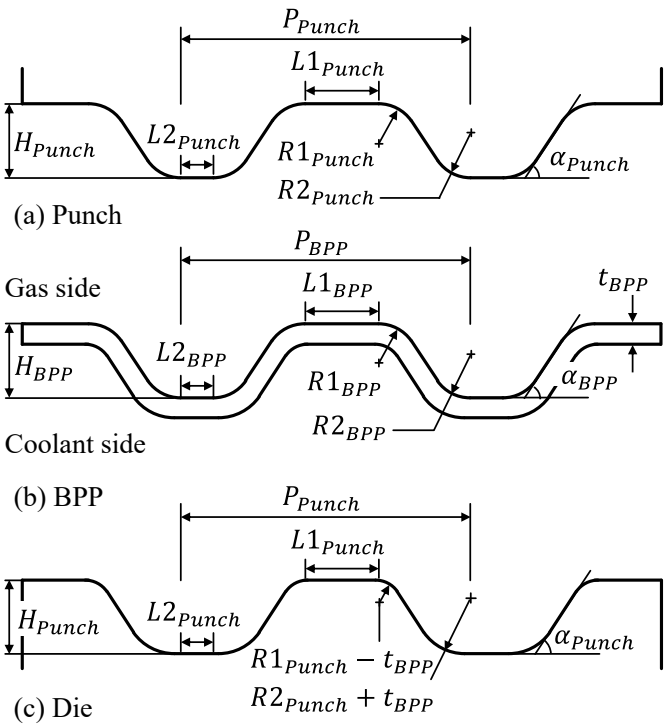


Fig. 1. Geometrical parameters.

An automated workflow was developed to integrate FE simulation, geometric measurement, and Bayesian optimization, as illustrated in Figure 2. After forming and springback simulation, the surface profile of the BPP was automatically extracted from the ABAQUS output database. An internally developed measurement algorithm was then used to determine the six geometric parameters of the simulated BPP. The mean relative error between the resulting and target geometries was calculated based on these six parameters and used as the objective function for optimization as shown in Equation (1).

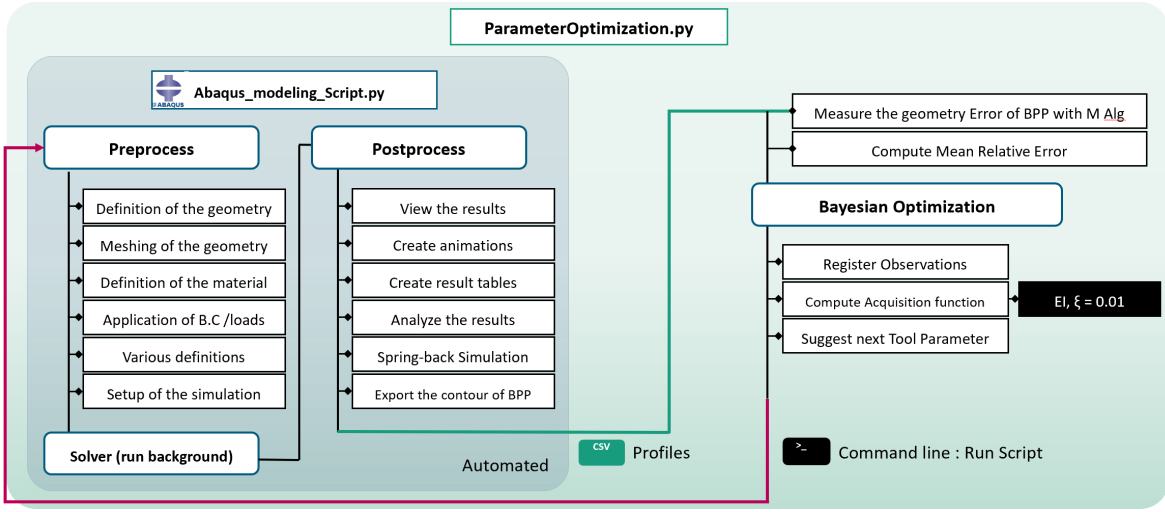


Fig. 2. Automated workflow for Bayesian optimization-based tool geometry compensation.

$$f(x) = \frac{1}{6}\sum_{i=1}^{6}\left|\frac{\hat{a}_{BPP,i} - a_{BPP,i}}{a_{BPP,i}}\right| \tag{1}$$

where, $a_{BPP,i} \in \{H_{BPP}, \alpha_{BPP}, R1_{BPP}, R2_{BPP}, L1_{BPP}, L2_{BPP}\}$ is a set of target geometric parameters of the BPP, $\hat{a}_{BPP,i} \in \{\widehat{H}_{BPP}, \hat{\alpha}_{BPP}, \widehat{R1}_{BPP}, \widehat{R2}_{BPP}, \widehat{L1}_{BPP}, \widehat{L2}_{BPP}\}$is the set of measured geometric parameters of the BPP and i denotes the parameter index.

Bayesian optimization was applied to determine the compensated punch geometry while reducing the number of required FE simulations. In Bayesian optimization, a GP-based surrogate model approximates the relationship between the input tool parameters and the geometric error based on the available observations. The GP provides both a predicted value and its uncertainty for unexplored parameter combinations. Based on this information, an acquisition function determines the next parameter set to evaluate by balancing exploitation of regions with promising predicted performance and exploration of regions with high uncertainty. In this study, EI was employed as the acquisition function.

Eight parameter sets obtained from a CCF design were used as the initial dataset. Based on the simulation results of these initial parameter sets, the GP surrogate model was constructed and EI suggested the next promising tool geometry. Each newly suggested parameter set was evaluated through FE simulation, and the resulting geometric error was added to the existing dataset to update the surrogate model. Candidate parameter sets that violated the predefined pitch constraint or parameter bounds were excluded from the optimization process. This simulation-measurement-optimization cycle was repeated automatically until the stopping criterion was satisfied.

The performance of the proposed framework was evaluated by comparison with the conventional DoE approach. As shown in Figure 3, Bayesian optimization identified the optimal parameter set after 10 simulations, whereas the DoE approach evaluated 43 predefined parameter sets. The minimum geometric error obtained by Bayesian optimization was 6.01%, which was slightly lower than the 6.39% obtained using DoE. More importantly, the number of required simulations was reduced by 76.74%. Since the computational time required for each FE simulation was approximately constant, the total simulation time was also reduced from 70.95 hours to 16.50 hours.

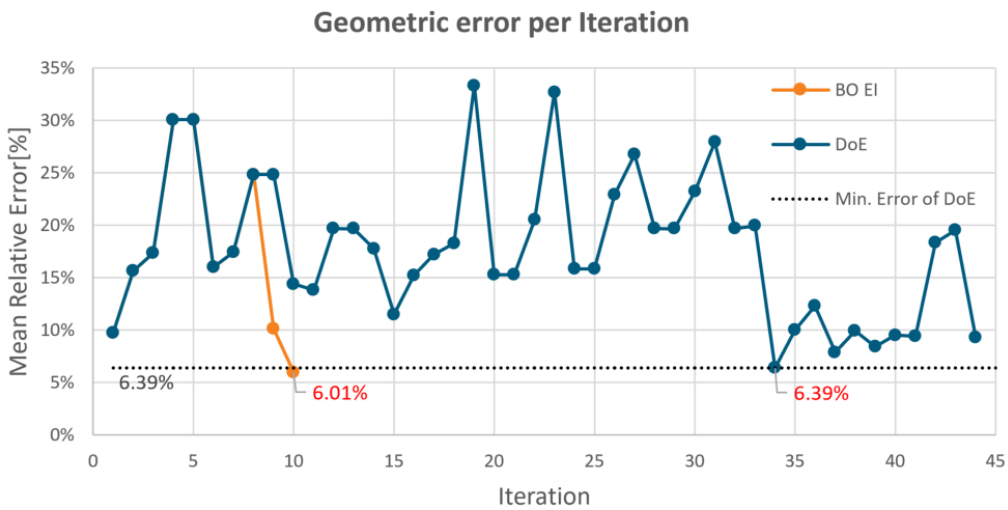


Fig. 3. Comparison of optimization progress between Bayesian optimization and DoE.

Table 1. Comparison of optimization performance.

Method	Minimum geometric error	Number of simulations	Total runtime
Bayesian optimization	6.01%	10	16.50 hours
DoE (CCF)	6.39%	43	70.95 hours

The results show that Bayesian optimization can efficiently search for promising tool geometries. In contrast to DoE, where parameter combinations are predefined before the simulations are conducted, Bayesian optimization sequentially updates the surrogate model and determines the next evaluation point based on the accumulated results. This adaptive search strategy enabled the optimization to achieve a lower geometric error with fewer simulations. In the near future, the optimized tool geometry will be validated using an in-house test bench.

3. Conclusions

This study proposed a Bayesian optimization-based framework for tool geometry compensation of metallic BPPs considering geometric deviations caused by springback. By integrating FE simulation, geometric measurement, and optimization into an automated workflow, compensated tool parameters were efficiently identified with minimal manual intervention. Compared with the conventional DoE approach, Bayesian optimization achieved a lower geometric error of 6.01% while reducing the number of required simulations from 43 to 10, corresponding to a 76.74% reduction in simulation iterations. These results demonstrate that Bayesian optimization can provide an efficient approach for FE-based tool geometry compensation, particularly for manufacturing processes where high geometric accuracy and computationally expensive simulations are involved.

References

1. L. Fan, T. Zhengkai, H. C. Siew, "Recent development of hydrogen and fuel cell technologies: A review", Energy Reports, vol. 7, pp. 8421–8439, 2021. https://doi.org/10.1016/j.egyr.2021.08.003
2. S. Karimi, N. Fraser, B. Roberts, F.R. Foulkes, "A Review of Metallic Bipolar Plates for Proton Exchange Membrane Fuel Cells: Materials and Fabrication Methods", Advances in Materials Science and Engineering, vol. 2012, pp. 1–10, 2012. https://doi.org/10.1155/2012/828070
3. J. Liu, L. Zhang, B. Yuan, Y. Zhang, Z. Yang, J. Huang, "Design and development of coating for metallic bipolar plates in proton exchange membrane fuel cell (PEMFC): A review", Materials & Design, vol. 246, pp. 1–6, 2024. https://doi.org/10.1016/j.matdes.2024.113338
4. X. Li, I. Sabir, "Review of bipolar plates in PEM fuel cells: Flow-field designs", International Journal of Hydrogen Energy, vol. 30, pp. 359–371, 2005. https://doi.org/10.1016/j.ijhydene.2004.09.019
5. M.T. Tran, D.H. Lee, H.W. Lee, D.K. Kim, "Formability improvement in multi-stage stamping of ultra-thin metallic bipolar plate for proton exchange membrane fuel cell", International Journal of Hydrogen Energy, vol. 47, pp. 40008–40025, 2022. https://doi.org/10.1016/j.ijhydene.2022.09.163
6. P. Zhang, M.P. Pereira, B.F. Rolfe, D.E. Wilkosz, P. Hodgson, M. Weiss, "Investigation of material failure in micro-stamping of metallic bipolar plates", Journal of Manufacturing Processes, vol. 73, pp. 54–66, 2022. https://doi.org/10.1016/j.jmapro.2021.10.044
7. A.R. Singh, T. Bashford-Rogers, D. Marnerides, K. Debattista, S. Hazra, "HDR image-based deep learning approach for automatic detection of split defects on sheet metal stamping parts", The International Journal of Advanced Manufacturing Technology, vol. 125, pp. 2393–2408, 2023. https://doi.org/10.1007/s00170-022-10763-6
8. C. Li, Z. Xu, L. Peng, X. Lai, "An electric-pulse-assisted stamping process towards springback suppression and precision fabrication of micro channels", International Journal of Mechanical Sciences, vol. 218, 107081, 2022. https://doi.org/10.1016/j.ijmecsci.2022.107081
9. S. Xu, K. Li, Y. Wei, W. Jiang, "Numerical investigation of formed residual stresses and the thickness of stainless steel bipolar plate in PEMFC", International Journal of Hydrogen Energy, vol. 41, pp. 6855–6863, 2016. http://dx.doi.org/10.1016/j.ijhydene.2016.03.005

Federated Data Ecosystems as an Enabler of the Circular Economy-Manufacturing-X

Ro, Seonhi[1], Schmidt, Lara[2], Dietrich, Katrin[3], Revellio, Ferdinand[4], Mügge, Janine[5] and Vollenbruch, Paul[6]

[1] Labs Network Industrie 4.0 e.V. Berlin, Germany, [2+3] Fraunhofer-Institut für Integrierte Schaltungen IIS, Erlangen, Germany,[4]Siemens AG, Berlin, [5] Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik IPK, Berlin Germany, [6]Siemens Mobility GmbH

seonhi.ro@hotmail.com

Abstract. It highlights the transformative role of Federated Data Ecosystems in supporting the implementation of a Circular Economy, particularly through the use case of R-Grading. Given the challenges companies face in transitioning from linear to circular business models, R-Grading offers data-driven decision support to identify the optimal R-strategies for products and components. By integrating lifecycle-data and promoting secure and decentralized data-exchange between different actors in the value-chain, Federated Data Ecosystems create the necessary transparency and interoperability. These systems enable companies to develop more efficient R-strategies executions with both environmental and economic benefits. The implementation of R-Grading in various industries - such as automotive, aviation, railway, machine and fluid power - is promising but still in its infancy. Recommendations for promoting the Circular Economy and Federated Data Ecosystems, aimed at product manufacturers, digital service providers, and legislators, underscore the need for a coordinated approach to overcome existing hurdles.
Overall, the combination of technological innovations and a structured, trust-based exchange of data is crucial to fully exploit the potential of the Circular Economy and thus enable a more sustainable future. By creating new service offerings and increasing service revenues, companies can diversify their business models and reduce their dependency on selling new products. Keeping materials in circulation not only improves resource availability and decreases material costs but also strengthens the resilience of supply chains, making organizations less vulnerable to disruptions. Ultimately, the collaborative use of R-Grading tools fosters more robust, sustainable, and adaptable value chains, benefiting both individual enterprises and the broader economy.

Keywords: Circular economy, R-Grading, Manufacturing-X, Federated Data Ecosystem, Sustainability, Catena-X, Factory-X, Fluid 4.0, Railway-X, Aerospace-X

This article is already published as Fraunhofer Gesellschaft Research results Federated Data Ecosystems as an Enabler of the Circular Economy (https://publica.fraunhofer.de/entities/publication/9e38532c-ce85-4e6b-84e2-cd3c7974f3cb) within the Manufacturing-X Projects funded by German Federal Ministry of Economy and Energy. Here shows the shortened version of that publication.

Rising costs, limited growth opportunities, and increasing dependence on expensive raw materials are driving companies to maximize the value and lifespan of existing assets through smarter maintenance, digital retrofits, and circular practices. At the same time, the traditional linear "take-make-dispose" economy contributes significantly to resource depletion, waste, climate change, and biodiversity loss, making the Circular Economy an attractive alternative that decouples growth from primary resource consumption.

However, implementing circular strategies such as reuse, repair, refurbishment, and remanufacturing is often hindered by limited access to reliable lifecycle information. Federated Data Ecosystems address this challenge by enabling secure and standardized data sharing across value chains, supporting data-driven R-Grading decisions that identify the most appropriate circular strategy and accelerate the adoption of Circular Economy business models.

1. Federated Data Ecosystems as enabler of a Circular Economy

1.1 Challenges in implementing a Circular Economy using the example of the "R-Grading" use case

- Insufficient use-phase information and lack of product knowledge
- Digitalized expert knowledge for product diagnostics

- Inefficient processes
- Interoperability
- End-of-life treatment is often not sufficiently considered at the beginning of the product life cycle

R-Grading is a data-driven decision support approach that combines product, lifecycle, operational, economic, and environmental data to determine the most suitable circular strategy for a given asset, such as reuse, repair, refurbishment, remanufacturing, or replacement. By improving the quality and efficiency of decision-making, it creates opportunities for more profitable and sustainable circular business models. As the required data originates from multiple stakeholders across the value chain, Federated Data Ecosystems provide the secure and trustworthy data exchange needed to enable effective R-Grading implementation.

1.2 Federated Data Ecosystems as the solution

Traditional production networks rely on rigid, often proprietary and centrally managed data exchange mechanisms that limit flexible collaboration, whereas emerging Federated Data Ecosystems enable secure, sovereign, and seamless cross-company data sharing across complex value chains, supporting use cases such as the Circular Economy (Otto et al., 2019).

- Data and Technology Sovereignty
- Interoperability
- Trust-by-Design

Federated Data Ecosystems provide a decentralized and interoperable framework for secure and sovereign cross-company data exchange, enabled by technical data exchange stacks such as EDC, AAS, OPC-UA, and the MX-Port concept, which are being developed and applied across initiatives including Catena-X, Factory-X, Fluid 4.0, Railway-X, and Aerospace-X to support collaborative use cases such as R-Grading and the Circular Economy.

2. R-Grading enabled by Federated Data Ecosystems

R-Grading is a key enabler of Circular Economy processes, using Federated Data Ecosystems to support data exchange and collaboration across value chain actors. Its implementation is assessed across industries based on the supported R-strategies; involved actors and their roles; decision criteria covering technical feasibility, profitability, and environmental aspects; data categories including technical, economic, environmental, operational, and assessment data; data models such as IDTA submodels, Catena Semantic Aspect Models (SAMM), and proprietary solutions; enabling technologies including AAS, EDC, MX-Port variants, and standards-based proprietary solutions; and the tool's maturity level measured using the Technology Readiness Level (TRL) according to ISO 16290.

2.1 Automotive (Catena-X)

The R-Grading tool of Catena-X focuses on the EoL-vehicle dismantling decision process. It enables faster, data-driven decision-making of decomissioned end-of-life-vehicles and while assisting the dismantling, it leads to the most suitable R-strategy for automotive components. The project has resulted in initial data models, a conceptual decision-making logic and demonstrator mock-up.

2.2 Aviation industry (Aerospace-X)

R-Grading in Aerospace-X supports end-of-life as well as maintenance decisions regarding the selection of the most suitable R-strategy for aviation components. The project will lead to initial data models and conceptual decision-making logic for exemplary components.

2.3 Railway industry (Railway-X)

The R-Grading tool of Railway-X focuses on an AI-based decision support for the end-of-life management, maintenance and repair work. During the operational work within the product's service life and at the end of a product's lifecycle, the most suitable R-strategy will be suggested by using knowledge and data. This will be tested on exemplary rail components.

2.4 Machine and plant engineering (Factory-X)

Factory-X focuses on creating a unified infrastructure for the exchange of R-Grading-related operational data between machine builder, factory operators, and other relevant actors during the machine's use phase or at end of its life. The specific R-Grading algorithm itself is not part of the solution, but is proprietary of the respective component manufacturers. The results will be shared back with the factory operator via the R-Grading tool.

2.5 Fluid power industry (Fluid 4.0)

R-Grading in Fluid 4.0 illustrates the process and decision logic for selecting the best R strategy for hydraulic or pneumatic products during and after their use. A data model has been developed that serves as the basis for providing information about the product, intensity of use, and possible R-strategies. This approach serves as a blueprint for instantiating product-specific R-strategy processes and logics.

2.6 Comparison of different industries

R-Grading tools across different industries share the goal of identifying the most appropriate R-strategy, but they differ in their focus areas, methodologies, and application contexts. All tools are currently at the conceptual stage and are being developed within publicly funded research projects, so practical performance cannot yet be evaluated. While all approaches focus on the use and post-use phases of assets and support strategies such as reuse, repair, refurbish, remanufacture, recycle, and recover, none address the pre-use strategies (refuse, rethink, reduce) and repurpose is largely absent. Industry-specific implementations vary, with Catena-X concentrating on end-of-life vehicles and closing-the-loop strategies, while Factory-X, Fluid 4.0, Aerospace-X, and Railway-X place greater emphasis on extending product lifecycles through maintenance, repair, refurbishment, and remanufacturing (see Fig. 1)

Fig. 1. Comparison of relevant R-strategies in the different industries

Fig. 2. Comparison of actors and roles in the R-Grading use case of R-Grading use case of different industries.

The selected R-strategies are not linked to specific R-Grading tool characteristics, indicating that the tools are generally adaptable across industries, although regulatory constraints may limit certain strategies for safety-critical components, particularly in the aerospace and automotive sectors. A common feature across all projects is the involvement of three key roles—data providers, infrastructure providers, and tool users—with actors often assuming multiple roles and contributing either static product data, dynamic lifecycle data, or data exchange infrastructure. Figure 2 illustrates that, despite minor differences such as value-chain actors also serving as infrastructure providers in Factory-X and Fluid 4.0, the roles, interactions, and responsibilities of stakeholders remain largely consistent across industries and projects.

Across all projects, technical feasibility is considered the most important criterion for selecting an R-strategy, followed by economic factors such as costs, process effort, and component availability, while environmental indicators (e.g., product carbon footprint) play a supporting role. The R-Grading tools rely on standardized data exchange using IDTA sub-models, Catena Semantic Aspect Models (SAMM), and proprietary models, covering static technical data, economic data, environmental data, dynamic operational data, and assessment data, with projects such as Factory-X and Fluid 4.0 employing data aggregation mechanisms to address data-sharing concerns. All solutions are implemented within Federated Data Ecosystems, using the AAS for data provision and technologies such as EDC, MX-Port, and standards-based proprietary solutions for data exchange. While all tools are currently below TRL 7, the projects contribute complementary capabilities—Catena-X lays the foundation, Aerospace-X adapts it to aviation, Factory-X focuses on infrastructure and business processes, Fluid 4.0 on data and actor interactions, and Railway-X on decision algorithms—collectively forming a holistic but still evolving R-Grading ecosystem.

3. Recommendations for enabling a Circular Economy with Federated Data Ecosystems

These Recommendations describe targeted measures designed to accelerate the implementation and operationalization of use cases such as R-Gading to facilitate the transition to a Circular Economy enabled by Federated Data Ecosystems.

Recommendations for product manufacturers are considered the following aspects:
- R-Grading readiness in product development and business models and
- From pilot projects to scaling effects and economic viability:

Recommendations for digital service providers:
- Simplified R-Grading Apps for user accessibility
- Formalization of expert knowledge for Circular Economy with appropriate AI methods

Recommendations for cross-company collaboration:
- Trust structures to overcome reluctance to share data
- Initiatives for adoption of existing standards

- Development of joint R-Grading tool kit

Recommendations for legislators:
- Financial support for the basic digitization of SMEs
- Create further (financial) incentives for Circular Economy practices
- Simplified legal guidelines for data sharing

The more details of recommendations described in the original publication (https://publica.fraunhofer.de/entities/publication/9e38532c-ce85-4e6b-84e2-cd3c7974f3cb).

4. Conclusions

The study demonstrates that Federated Data Ecosystems are a key enabler of the Circular Economy, as they provide the secure, decentralized, and interoperable data exchange needed for R-Grading, a data-driven approach that identifies the most appropriate circular strategy for products, components, and materials. By combining lifecycle data from multiple stakeholders, R-Grading supports more effective implementation of circular practices while generating both environmental and economic benefits. Although current implementations across industries such as automotive, aviation, railway, machine engineering, and fluid power are still in an early development stage, they show strong potential to improve resource efficiency, supply-chain resilience, and new service-based business models. Overall, the article concludes that a combination of technological innovation, trusted data sharing, and cross-industry collaboration is essential for scaling Circular Economy solutions and creating more sustainable and adaptable value chains.

References

1. Alcayaga, A. & Hansen, E. G. (2025). Smart circular economy as a service business model: an activity system framework and research agenda. R&D Management, 55(2), 508-530. https://doi.org/10.1111/radm.12707
2. Bühler, L., SCHUSTER, T. & PFLAUM, A. (2025). Twin Transformation: Insights from an Umbrella Review. International Journal of Innovation Management, 29(05n06), Artikel 2540007. https://doi.org/10.1142/S1363919625400079
3. Ellen MacArthur Foundation. (2013). Towards the circular economy Vol. 1: an economic and business rationale for an accelerated transition. https://www.ellenmacarthurfoundation.org/towards-the-circular-economy-vol-1-aneconomic-and-business-rationale-for-an
4. European Commission. (2020). Circular economy action plan – For a cleaner and more competitive Europe. Publications Office of the European Union. https://doi.org/10.2779/05068
5. Otto, B. (2022). A federated infrastructure for European data spaces. Communications of the ACM, 65(4), 44–45. https://doi.org/10.1145/3512341
6. Otto, B., Steinbuss, S., Teuscher, A., Lohmann, S., Bader, S., Birnstil, P., Böhmer, M., Brost, G., Cirullies, J., Eitel, A., Ernst, T., Geisler, S., Gelhaar, J., Gude, R., Haas, C., Huber, M., Jung, C., Jürjens, J., Lange, C., et al. (2019). Reference Architecture Model. Version 3.0. https://doi.org/10.24406/publica-fhg-299836
7. Potting, J., Hekkert, M. P., Worrell, E. &Hanemaaijer, A. (2017). Circular Economy: Measuring Innovation in the Product Chain. Planbureauvoor de Leefomgeving: Bd. 2544. PBL Publishers.

A Proposal on the Disciplined Use of Large Language Model (LLM) AI for Engineering Research and Code Development

Junsu Shin[1]

[1] Korea National University of Transportation, Chungju, Republic of Korea
junsu.shin@ut.ac.kr

Abstract. Large language models (LLMs) have moved rapidly from text assistants to practical producers of research software, and a growing share of the computational apparatus of engineering research is now drafted with their help. This paper argues that the decisive variable in that workflow is not the capability of the model but the *quality of the dialogue* that precedes the code. Because an LLM generates by sequentially conditioning on the accumulated prompt–response trace, a shallow opening exchange fixes an impoverished context that later correction does not fully repair. Three failure modes of AI-generated research code are identified — silent physical error, unstated assumption and confident interpolation — none of which execution-based testing reliably detects. Drawing on the harness-engineering, Research–Plan–Implement (RPI) and multi-agent test-driven-development patterns discussed in the 8th Light session *AI in the Open: Turning LLMs into Reliable Engineering Partners* [1], a three-part discipline is proposed: structure the dialogue before requesting code, wrap generation in a verification harness, and recognise that the competences governing the first are linguistic rather than purely technical.

Keywords: Large Language Models, Engineering Software, Scientific Code Verification, Agentic Workflow, Human–AI Collaboration.

1. Introduction

The adoption of LLM-based assistants in engineering research has followed an unusually steep curve. What began as a convenience for drafting documentation is now a routine part of producing the computational apparatus of research itself: pre-processors, post-processing pipelines, parameter sweeps, surrogate models and the glue code binding commercial solvers to bespoke analysis. Two capabilities drive this: the growth of usable context, which lets a model hold a solver interface, a material model and the governing equations at once; and tool use, which lets assistants run their own code, read the traceback and revise.

The unresolved question is a different one. Research code is not merely software that runs; it is an instrument whose output enters the scientific record. A wrong sign convention or a boundary condition on the wrong face does not produce a crash — it produces a plausible number, and LLMs are by construction optimised for plausible output. The question is not whether to use LLMs in engineering research, but under what discipline their use is defensible.

2. The Verification Gap

Three failure modes deserve naming. Silent physical error: the code is numerically sound but encodes the wrong physics. Unstated assumption: the model selects a default — an integration scheme, a tolerance, a linearisation — reasonable in general and wrong for the regime under study. Confident interpolation: when a request is underspecified the model does not stop and ask, but fills the gap with the most likely continuation, which is by definition the conventional case rather than the researcher's.

None of these is removed by testing alone, because the test the researcher writes is generated from the same misunderstanding as the code; the most damaging failures are those least likely to announce themselves at run time. Verification must reach back past the code, to the specification — and past that, to the conversation in which it was formed.

3. Sequential Conditioning: Why Dialogue Precedes Direction

An LLM does not answer a question; it continues a document. Every token it emits is conditioned on the whole of the preceding exchange. Two properties follow. First, the process is path-dependent: the context established in the opening turns is not a preamble that is later discarded but the substrate on which all subsequent reasoning is built, so an incorrect early frame — an implicit assumption of steady state, of linearity, of an idealised boundary — is carried forward, and later corrections layer on top of it rather than replacing it. Second, it is self-reinforcing: because the model conditions on its own prior output, an early plausible-but-wrong statement raises the probability of later statements consistent with it. Two conversations beginning from the same question therefore diverge sharply depending on how the first few turns are spent (Fig. 1).

The practical implication inverts the usual instinct, which is to ask for the code and then correct it. The discipline proposed here is to spend the early turns establishing shared ground instead: the governing equations and their domain of validity, the regime and its dimensionless parameters, what is fixed and what is varied — then asking the model to restate the problem in its own words and enumerate its assumptions before a line is written. Disagreement surfaced here is cheap; the same disagreement after a parameter study has run is not.

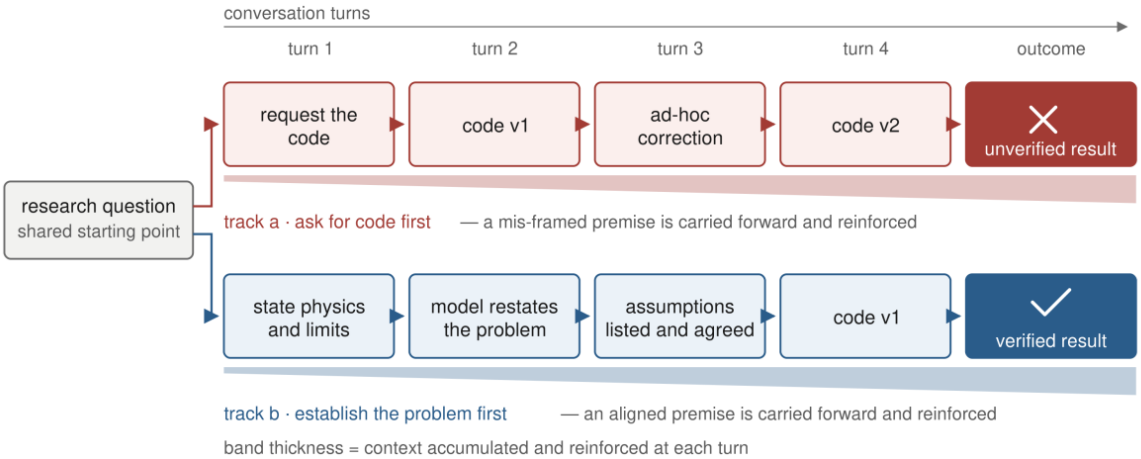


Fig. 1. Divergence of two interaction trajectories from a common research question. The widening bands show context accumulating and being reinforced at every turn, which is why the cost of correcting a mis-framed premise rises with each turn of delay.

4. Lessons from Agentic Engineering Practice

The argument converges with independent industrial practice. In the 8th Light session *AI in the Open: Turning LLMs into Reliable Engineering Partners* (19 May 2026), Travis Frisinger and Eric Olson describe workflows intended, in their words, to bring software craft to working with LLMs and to turn AI from a code generator into a reliable engineering partner [1]. Three of their ideas transfer directly.

Harness engineering. The unit of design is not the prompt but the harness surrounding the model — context, tools, enforced verification steps and orchestration logic [1]. For a research group this reframes the deliverable: what should be maintained is not a library of clever prompts but a reusable scaffold that automatically confronts any generated routine with unit tests, dimensional checks, limit cases and regression against a trusted baseline.

Structured RPI loops. The Research–Plan–Implement pattern [1,4] separates the workflow into three phases with a human checkpoint between each. Its value here is that it makes the specification an inspectable artefact: the plan can be read by a supervisor, disputed by a co-author and archived with the results — precisely what Section 2 demands.

Multi-agent TDD and human-in-the-loop. Separating the agent that writes the test from the agent that writes the implementation breaks the shared-misunderstanding failure mode, because the two artefacts no longer originate from a single chain of reasoning; the human acts as arbiter at each boundary. Frisinger's writing on “vibe coding” [2] and on patterns across large corpora of user conversations [3] reinforces the same conclusion from the opposite direction: unstructured, momentum-driven interaction is exactly the condition under which unexamined assumptions propagate. Fig. 2 assembles these elements into one workflow.

5. The Humanistic Requirement

If the quality of engineering output depends on the quality of the preceding dialogue, then the competences governing that dialogue become engineering competences — though not the ones engineering curricula normally cultivate. Such an exchange requires the ability to make tacit knowledge explicit; precision of language, since the model responds to what was written rather than what was meant; the discipline of interrogation, asking the model to justify a choice and to state what would falsify it; and the self-awareness to notice when a fluent answer produces agreement rather than understanding. These are skills of reading, writing, argument and critical scepticism, and the claim here is deliberately strong: the humanistic education that engineering programmes treat as peripheral is becoming a determinant of technical quality.

6. Recommendations

(i) Never request code in the opening exchange; establish the physical problem, governing equations, regime and validity limits first, and require the model to restate them and enumerate its assumptions. (ii) Adopt an explicit phase separation — Research, Plan, Implement — with a human decision point at each boundary, and version the plan alongside the code. (iii) Build a verification harness once per group, not per project: dimensional and conservation checks, analytical limit cases, regression against a trusted reference. (iv) Separate the origin of tests from that of implementation. (v) Preserve the conversation transcript as part of the research record; it is the specification, and currently the least reproducible element of AI-assisted computational work.

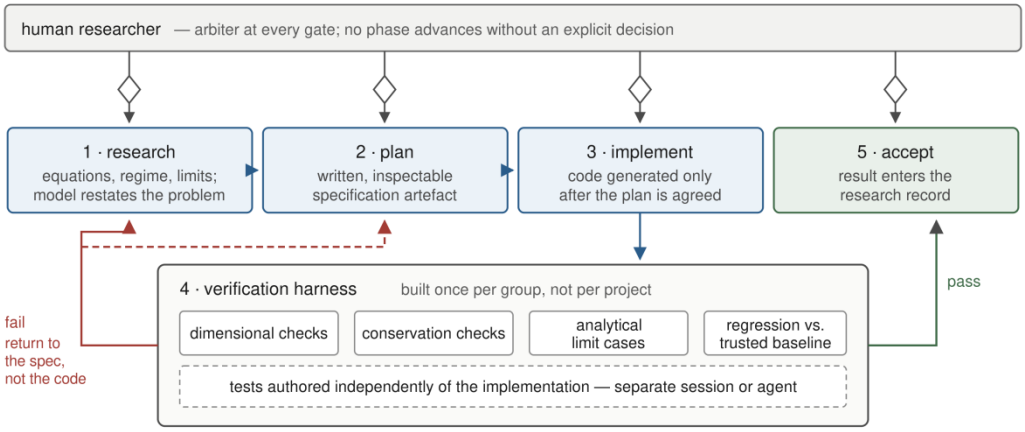


Fig. 2. Proposed workflow: a phase-separated Research–Plan–Implement loop with a persistent verification harness and a human decision point at each boundary. A harness failure returns to the specification, not to the code.

7. Conclusions

LLMs have become useful producers of research code; whether to use them is settled in practice. What is not settled is the discipline of that use. Because these models generate by sequential conditioning on an accumulated context, the decisive intervention occurs early — in the structure and depth of the dialogue that establishes the problem — not late, in the review of generated output. Industrial practice has reached the same conclusion through harnesses, phase-separated loops and human checkpoints. Engineering research should adopt the same posture, and treat the competences underwriting it as part of an engineer's education.

References

1. Frisinger, T., Olson, E.: AI in the Open: Turning LLMs into Reliable Engineering Partners. 8th Light Inc., YouTube (2026), https://www.youtube.com/watch?v=iTgqKTqtnow
2. Frisinger, T.: AI-Assisted Coding and the Rise of "Vibe Coding". 8th Light (2025), https://8thlight.com/insights/ai-assisted-coding-rise-of-vibe-coding
3. Frisinger, T.: What I Discovered After Analyzing 2,500 ChatGPT Conversations. 8th Light (2025), https://8thlight.com/insights/analysis-2500-chatgpt-conversations
4. Research, Plan, Implement (RPI). Agentic Engineering patterns, https://path.kilo.ai/introduction/patterns/rpi/

Material Division

Aims

The Material Division invites voluntary contributions from VeKNI members that reflect the broad and rapidly evolving research landscape across the materials sciences. Members are encouraged to share recent scientific advances, summaries of ongoing research projects, emerging technology trends, translational applications, industrial case studies, or perspectives on future directions and challenges in materials research

Scope

- **Synthesis & Fabrication**: Advanced synthesis routes, Chemical reaction engineering, Organic/Inorganic synthesis, and Novel composite materials

- **Advanced Characterization**: Structural, Chemical, Mechanical, Electrical, Thermal and Electrochemical property analysis using advanced techniques

- **Functional Applications**: High-performance structural materials, Energy storage/conversion devices, Electronic components, and Biomaterials

- **Data-Driven Material Science**: Processing Automation, Multi-scale simulation, Parameter optimization, and Machine learning in material design

Mechanism-Guided Selectivity Control in Metal-Nitrenoid Transfer

Suhyeon Kim[1]

[1]Institute of Organic Chemistry, RWTH Aachen University, Aachen, Germany
suhyeon.kim@rwth-aachen.de

Abstract. The precise control of reactivity and selectivity remains a central challenge in metal-nitrenoid transfer, where highly reactive intermediates can access multiple competing pathways. Catalyst and ligand design have therefore played a key role in controlling chemo-, site-, diastereo-, and enantioselectivity. In addition to conventional steric and electronic tuning, mechanistic understanding can provide a complementary strategy by identifying the origin of an undesired pathway and suggesting how the electronic structure of a metal nitrenoid may be altered to redirect reactivity. This short overview highlights two representative examples of such mechanism-guided selectivity control. In an iridium system, frontier molecular orbital analysis enabled orbital-symmetry tuning of an Irnitrenoid, switching alkene difunctionalization from a ligand-participating concerted pathway to a catalytic stepwise process and allowing highly stereoselective γ-lactam synthesis. In a copper system, open-shell nitrenoid reactivity was exploited to replace the usual closed-shell C–H insertion pathway with stepwise hydrogen-atom abstraction and radical rebound, enabling regio- and enantioselective δ-C–H amidation. Together, these studies illustrate how reaction-mechanism analysis can be translated into catalyst design for otherwise difficult selectivity control.

Keywords: Metal nitrenoid, selectivity control, asymmetric catalysis, reaction mechanism, catalyst design.

1. Introduction

Metal-nitrenoid transfer has emerged as a powerful platform for direct C–N bond formation from alkenes and C–H bonds. Because nitrogen-containing motifs are widespread in pharmaceuticals, natural products, and other bioactive molecules, the development of catalytic nitrene-transfer reactions with predictable selectivity has attracted sustained interest [1]. The challenge lies in the intrinsically high reactivity of nitrene-derived intermediates: depending on the metal, ligand, and substrate, several mechanistic pathways may be energetically accessible, giving rise to competing chemo-, site-, diastereo-, or enantioselective outcomes.

Considerable progress has been achieved by tailoring the primary and secondary coordination environments of the catalyst. For example, iridium catalysts have enabled selective γ-lactam formation by C(sp^3)–H amidation [2], while ligand modulation and non-covalent interactions have been used to alter chemoselectivity and site selectivity [3,4]. Chiral ligands can further impose high levels of enantiocontrol [5]. Beyond these approaches, detailed mechanistic analysis offers an additional design principle: if the pathway responsible for poor reactivity or selectivity can be identified, the electronic structure of the catalyst may be redesigned to favor a different pathway. The two studies discussed below illustrate this concept through complementary iridium- and copper-nitrenoid systems, in which control over the preferred mechanistic pathway directly enabled new selectivity profiles.

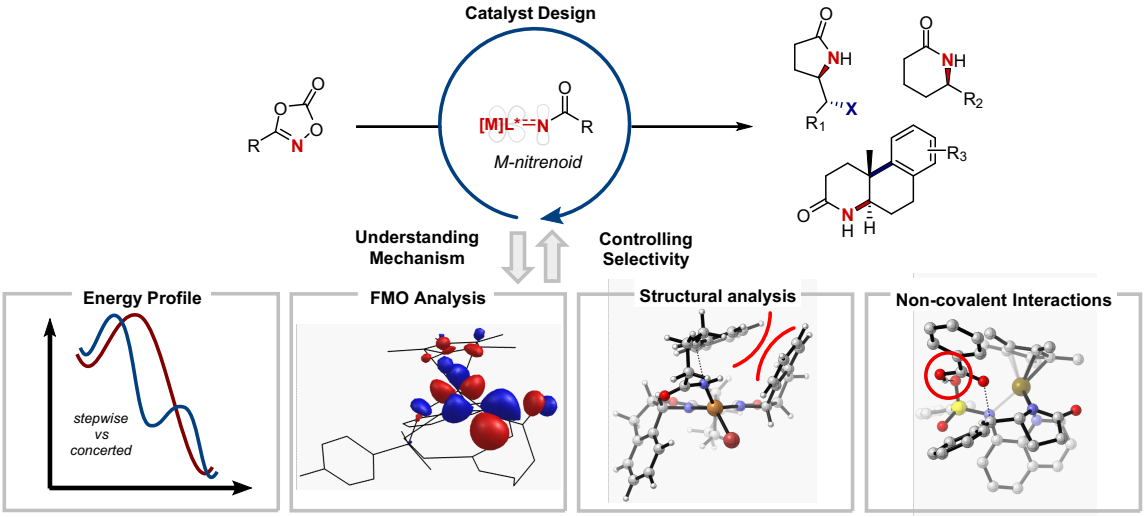


Fig. 1. Mechanism-guided catalyst design for selectivity control in metal-nitrenoid transfer.

2. Mechanism-Guided Control of Metal-Nitrenoid Reactivity

2.1 Orbital-Symmetry Tuning of Iridium Nitrenoids

Alkene difunctionalization provides rapid access to molecular complexity by forming two bonds across a C=C bond in a single operation. In acyl-nitrenoid chemistry, however, catalytic difunctionalization can be limited by competing reaction pathways. Previous studies showed that Cp*Ir complexes bearing bidentate ligands generate electrophilic Ir-nitrenoids with a three-centered LUMO that undergo concerted [3+2] cycloaddition with tethered alkenes [6]. Although this reactivity was highly selective, ligand participation produced a stable tridentate species and prevented efficient catalytic turnover.

Mechanistic and frontier molecular orbital analyses indicated that the concerted pathway was closely associated with the symmetry of the nitrenoid LUMO. Ligand modification was therefore used to disrupt the orbital interactions required for ligand participation and redirect the reaction toward a stepwise pathway. Sulfone-containing ligands proved particularly effective, enabling catalytic access to carbocation- or aziridine-type intermediates depending on the alkene substitution pattern [7]. These intermediates could be trapped by external nucleophiles, including carboxylic acids, alcohols, water, azide, and carbon nucleophiles. Extension of the same design principle to a chiral ligand framework allowed the simultaneous formation of two stereogenic centers in γ-lactams with excellent diastereo- and enantioselectivity, reaching >99:1 er. In this case, the improvement in reactivity and stereoselectivity arose from changing the preferred mechanistic pathway rather than simply enhancing the rate of the original concerted process.

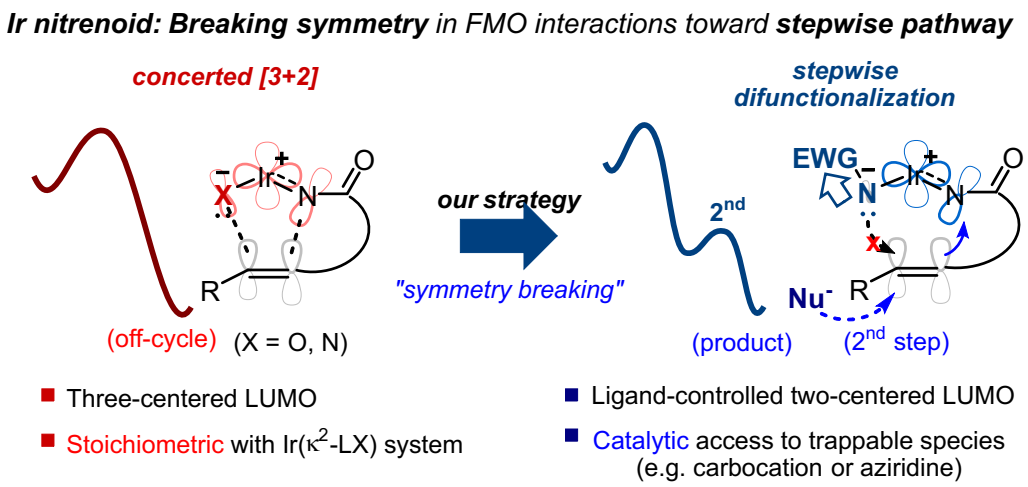


Fig. 2. Orbital-Symmetry Tuning of Iridium-Nitrenoid Reactivity.

2.2 Open-Shell Copper Nitrenoids for Regio- and Enantioselective C–H Amidation

A complementary mechanistic strategy was applied to the selective synthesis of chiral δ-lactams. Conventional closed-shell acylnitrenoids, including singlet Ir and Ru species, generally undergo concerted C(sp^3)–H insertion and prefer pathways leading to five-membered γ-lactams. As a result, selective δ-C–H amidation is challenging, especially when γ- and δ-C–H bonds compete within the same substrate. An alternative pathway based on open-shell nitrenoid reactivity was therefore considered to overcome this intrinsic preference.

A chiral Cu(I)/bisoxazoline catalyst was found to promote intramolecular δ-C–H amidation of dioxazolones with high regio- and enantioselectivity [8]. Experimental mechanistic probes and DFT calculations supported a stepwise open-shell pathway involving hydrogen-atom abstraction (HAT) followed by radical rebound. Regioselective HAT generates a carbon-centered radical at the δ position, while the subsequent radical-rebound step forms the C–N bond within the chiral catalyst environment and determines enantioselectivity. This mechanistic switch enabled selective functionalization of benzylic, propargylic, allylic, and tertiary δ-C–H bonds even in the presence of competing γ-C–H sites, and provided chiral δ-lactams with high regioselectivity and up to >99:1 er. Thus, open-shell reactivity was not merely tolerated as an intermediate feature but was deliberately used to redefine both site- and stereoselectivity.

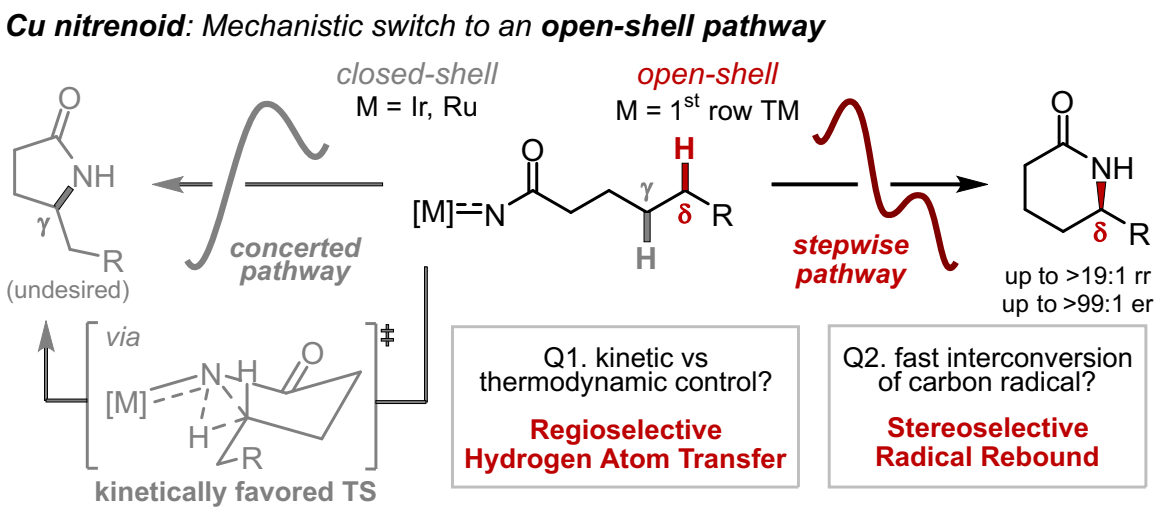


Fig. 3. Open-Shell Copper-Nitrenoid Reactivity for Selective C–H Amidation.

3. Conclusions

The two studies described here demonstrate complementary ways in which mechanistic understanding can guide selectivity control in metal-nitrenoid chemistry. In the iridium system, orbital analysis revealed the origin of an unproductive concerted pathway and suggested ligand modifications that enabled a catalytic stepwise process with high stereocontrol. In the copper system, changing the electronic character of the nitrenoid opened a radical HAT/rebound pathway that overcame the intrinsic preference for γ-C–H insertion and enabled regio- and enantioselective δ-lactam synthesis. These examples illustrate a broader principle in catalyst development: controlling which mechanistic pathway is accessible can be as important as optimizing the rate of a given elementary step, and can provide a useful route to otherwise difficult selectivity profiles.

References

1. Ju, M.; Schomaker, J. M.: Nitrene transfer catalysts for enantioselective C–N bond formation. Nat. Rev. Chem. 5, 580–594 (2021).
2. Hong, S. Y.; Park, Y.; Hwang, Y.; Kim, Y. B.; Baik, M.-H.; Chang, S.: Selective formation of γ-lactams via C–H amidation enabled by tailored iridium catalysts. Science 359, 1016–1021 (2018).
3. Hwang, Y.; Jung, H.; Lee, E.; Kim, D.; Chang, S.: Quantitative analysis on two-point ligand modulation of iridium catalysts for chemodivergent C–H amidation. J. Am. Chem. Soc. 142, 8880–8889 (2020).
4. Jung, H.; Schrader, M.; Kim, D.; Baik, M.-H.; Park, Y.; Chang, S.: Harnessing secondary coordination sphere interactions that enable the selective amidation of benzylic C–H bonds. J. Am. Chem. Soc. 141, 15356–15366 (2019).
5. Park, Y.; Chang, S.: Asymmetric formation of γ-lactams via C–H amidation enabled by chiral hydrogen-bond-donor catalysts. Nat. Catal. 2, 219–227 (2019).
6. Hong, S. Y.; Chang, S.: Stereodefined access to lactams via olefin difunctionalization: iridium nitrenoids as a motif of LUMO-controlled dipoles. J. Am. Chem. Soc. 141, 10399–10408 (2019).
7. Kim, S.; Kim, D.; Hong, S. Y.; Chang, S.: Tuning orbital symmetry of iridium nitrenoid enables catalytic diastereo- and enantioselective alkene difunctionalizations. J. Am. Chem. Soc. 143, 3993–4004 (2021).
8. Kim, S.; Song, S. L.; Zhang, J.; Kim, D.; Hong, S.; Chang, S.: Regio- and enantioselective catalytic δ-C–H amidation of dioxazolones enabled by open-shell copper-nitrenoid transfer. J. Am. Chem. Soc. 145, 16238–16248 (2023).

Orthogonal B/Si Activation of α-Silyl Allylboronic Esters for Stereodivergent Allyl–Allyl Coupling

Yongsuk Jung[1]

[1]Institute of Organic Chemistry, RWTH Aachen University, Landoltweg 1, 52074 Aachen, Germany
yongsuk.jung@rwth-aachen.de

Abstract. α-Silyl allylboronic esters are allyl-*gem*-dimetalloids bearing two selectively addressable carbon–metalloid bonds. We developed two complementary iridium-catalyzed allyl–allyl coupling pathways in which either the C–Si or C–B bond is selectively cleaved according to substrate substitution and activation mode. With γ-monosubstituted substrates, the β-silicon effect promotes S_E2'-type desilylation, affording (*E*)-1-boryl-substituted 1,5-dienes bearing vicinal tertiary stereocenters. Chirality pairing between the enantioenriched nucleophile and a chiral Ir-π-allyl intermediate provides access to all four stereoisomers with high chemo-, diastereo-, enantio-, and *E*/*Z* selectivity. In γ,γ-disubstituted substrates, steric congestion suppresses the desilylative pathway. Lewis-base activation of boron instead switches the reaction to deborylative coupling, furnishing (*E*)-1-silyl-substituted 1,5-dienes containing adjacent tertiary and all-carbon quaternary stereocenters. This pathway combines stereospecific chirality transfer with asymmetric iridium catalysis while retaining the silicon substituent. Together, these complementary reactions establish α-silyl allylboronic esters as a modular platform for chemoselective and stereodivergent C–C bond formation.

Keywords: allyl-*gem*-dimetalloids, iridium catalysis, stereodivergent synthesis, chemoselectivity, chirality pairing.

1. Introduction

Controlling chemo-, regio-, diastereo-, and enantioselectivity within a single carbon–carbon bond-forming reaction is a central challenge in modern synthesis [1]. The problem becomes especially demanding for polyfunctional nucleophiles, because several reactive sites may compete under catalytic conditions. Geminally functionalized organoboron and organosilicon compounds are attractive in this context: boron and silicon can each serve as a selectively addressable reaction handle, while the remaining group can be retained for subsequent derivatization. α–Silyl allylboronic esters are therefore not simply allylation reagents, but potentially programmable nucleophiles [2]. Enantioenriched members of this class can be obtained by kinetic resolution and used in stereospecific bond-forming reactions [3].

Our work established two complementary iridium-catalyzed allyl–allyl coupling manifolds. The first selectively engages silicon through β-silicon effect-guided desilylation, whereas the second activates boron through Lewis-base-induced deborylation. In both cases, chirality pairing between the enantioenriched nucleophile and a chiral Ir-π-allyl complex enables stereodivergent access to all four possible stereoisomers (Fig. 1).

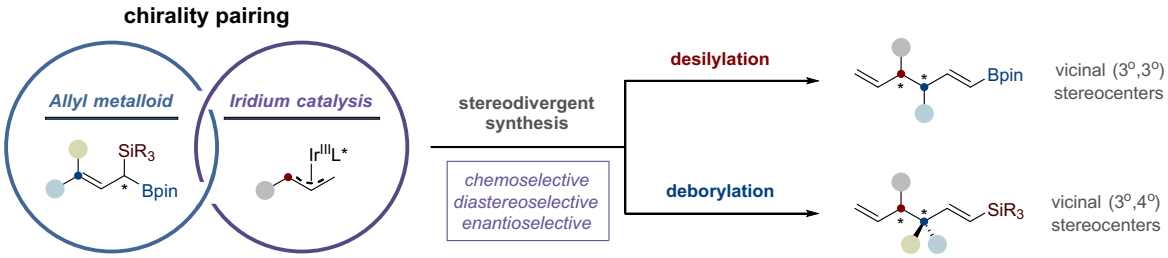


Fig. 1. Stereodivergent allyl–allyl coupling via chirality pairing of α-silyl allylboronic esters and iridium catalysts.

2. Experiments and Discussion

The geminal boron and silicon substituents of α-silyl allylboronic esters provide two distinguishable sites for bond activation. We examined how substrate substitution and activation mode determine which carbon–

element bond participates in iridium-catalyzed allyl–allyl coupling. Two complementary pathways were identified: a desilylative pathway governed by the β-silicon effect and a deborylative pathway enabled by Lewis-base activation of boron.

2.1 Desilylative Allyl–Allyl Coupling

The desilylative pathway was developed for γ-monosubstituted α-silyl allylboronic esters [4]. Under chiral iridium catalysis, branched allylic alcohols generate electrophilic Ir-π-allyl intermediates. Nucleophilic addition then proceeds through an S_E2'-type pathway in which the adjacent silicon group stabilizes developing positive charge at the β-position. This β-silicon effect accelerates selective C–Si cleavage without requiring an external activator, while the C–B bond remains intact. The products are therefore (*E*)-1-boryl-substituted 1,5-dienes that combine two newly controlled stereogenic centers with a synthetically useful alkenylboronic ester.

Most previous examples of stereodivergent catalysis relied on dual-catalyst systems, primarily with carbonyl nucleophiles [5]. In contrast, this approach uses a single iridium catalyst and chirality pairing with a non-carbonyl nucleophile to control the configurations of both newly formed stereocenters.

Stereochemical control arises from chirality pairing. Independently varying the configuration of the α-silyl allylboronic ester and the chiral ligand on iridium changes the matched and mismatched transition states. All four stereoisomers were obtained in good yields, generally with greater than 20:1 diastereomeric ratio, greater than 99:1 enantiomeric ratio, and greater than 99:1 *E*/*Z* selectivity in representative cases (Fig. 2). Mechanistic calculations supported the role of the β-silicon effect and explained how substrate and catalyst chirality cooperate during C–C bond formation. The retained boronic ester also enabled stereospecific downstream derivatization.

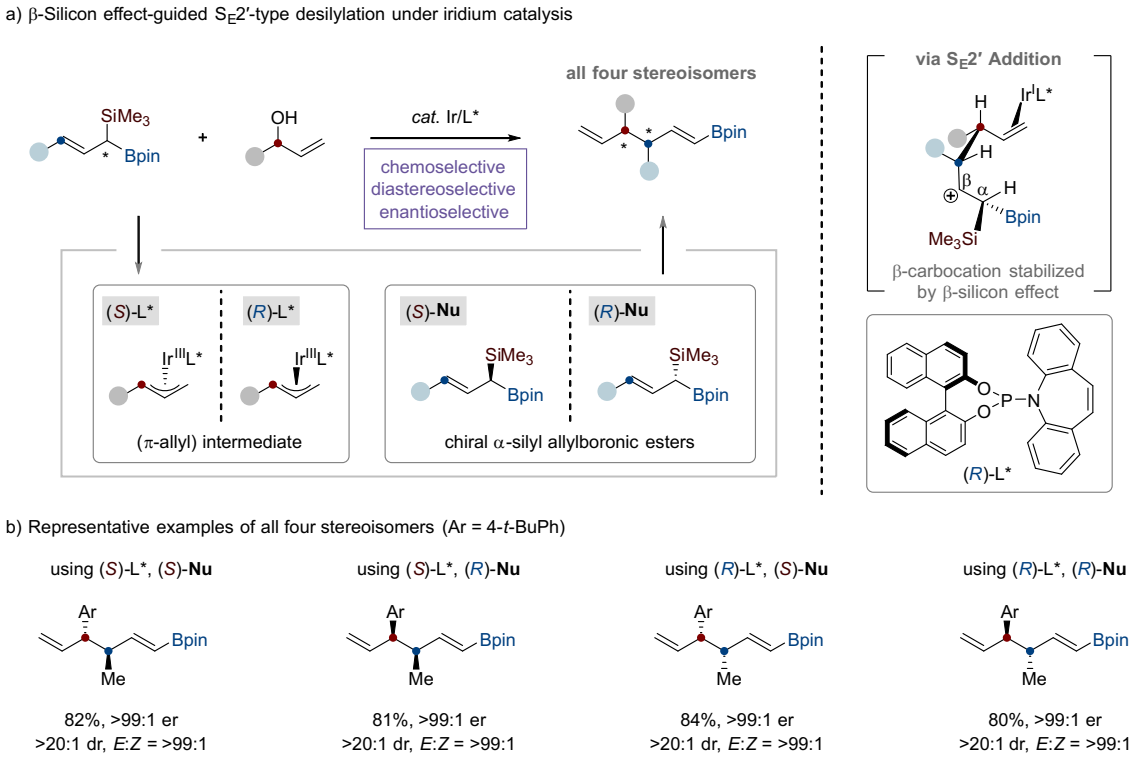


Fig. 2. Iridium-catalyzed allyl–allyl coupling utilizing the β-silicon effect: stereoselective synthesis of vicinal tertiary stereocenters.

2.2 Deborylative Allyl–Allyl Coupling

Increasing substitution at the γ-position changes the preferred reaction pathway. In γ,γ-disubstituted α-silyl allylboronic esters, steric congestion lowers nucleophilicity and makes the desilylative S_E2' process inefficient. This limitation was converted into a chemoselective switch by activating boron with an aryllithium reagent [6].

Formation of a boronate complex promotes deborylative allyl–allyl coupling while leaving the silicon substituent untouched. The reaction thus accesses sterically congested (*E*)-1-silyl-substituted 1,5-dienes bearing adjacent tertiary and all-carbon quaternary stereocenters.

This manifold combines stereospecific transfer from the enantioenriched allylboronic ester with asymmetric induction from the chiral iridium catalyst. Permuting nucleophile and ligand configurations again provide all four stereoisomers, with representative products formed in 57–67% yield, 13:1–15:1 diastereomeric ratio, 98:2–99:1 enantiomeric ratio, and greater than 99:1 *E*/*Z* selectivity (Fig. 3). DFT analysis was consistent with an asynchronous concerted transition state involving sequential C–B bond cleavage and C–C bond formation. The preserved silyl group also enables further functional-group interconversions, complementing the synthetic utility of the borylated products from the first pathway.

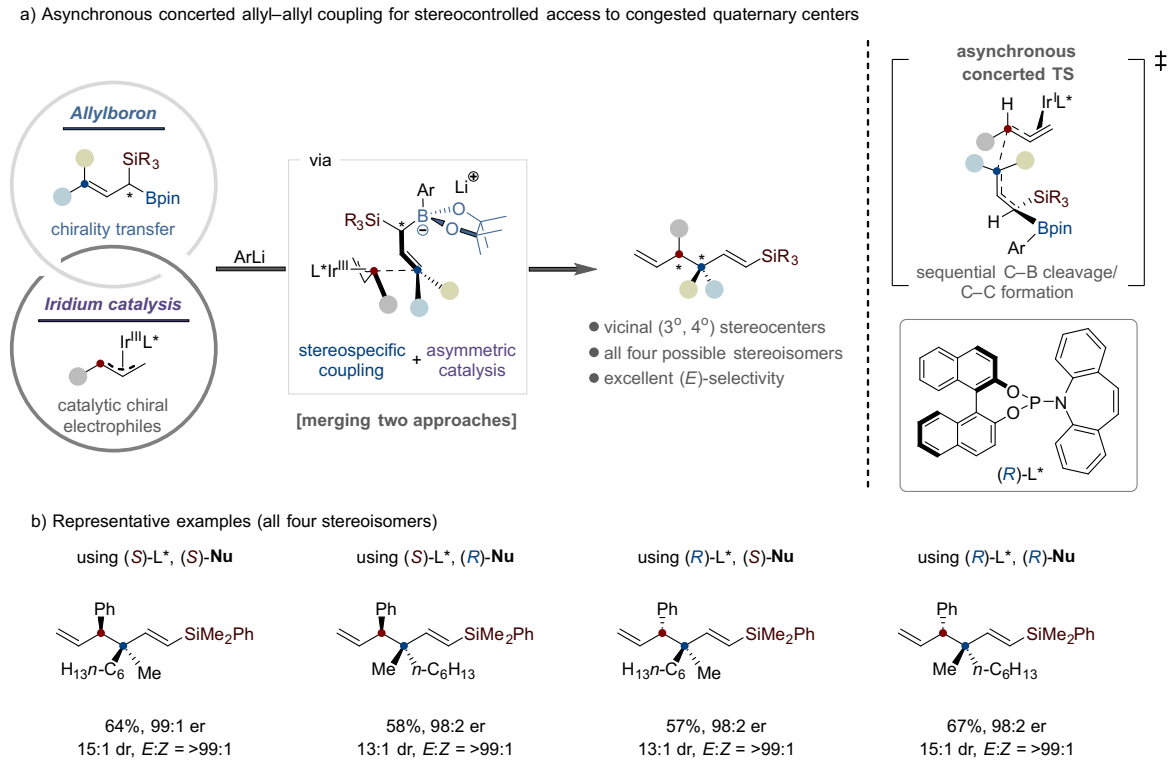


Fig. 3. Lewis-base-activated allyl–allyl coupling: stereocontrolled access to sterically congested all-carbon frameworks.

3. Conclusions

α-Silyl allylboronic esters illustrate how multifunctional reagents can be programmed through orthogonal activation. β-Silicon effect-guided desilylation efficiently constructs vicinal tertiary stereocenters while retaining boron, whereas Lewis-base-activated deborylation overcomes steric congestion to construct tertiary–quaternary stereocenters while retaining silicon. In both pathways, chirality pairing converts substrate and catalyst configuration into predictable access to all four stereoisomers. More broadly, this strategy shows that selective activation of different elements within a single reagent framework can expand both accessible chemical space and the level of stereochemical control available in catalytic synthesis.

References

1. (a) Cherney, A. H.; Kadunce, N. T.; Reisman, S. E. *Chem. Rev.***2015**, *115*, 9587–9652. (b) Ma, X.; Murray, B.; Biscoe, M. R. *Nat. Rev. Chem.***2020**, *4*, 584–599.
2. (a) Zhang, C.; Hu, W.; Morken, J. P. *ACS Catal.***2021**, *11*, 10660–10680. (b) Li, X.; Song, Q. *Org. Chem. Front.***2025**, *12*, 3129–3142.
3. Park, J.; Jung, Y.; Kim, J.; Lee, E.; Lee, S. Y.; Cho, S. H. *Adv. Synth. Catal.***2021**, *363*, 2371–2376.
4. Jung, Y.; Yoo, S. Y.; Jin, Y.; You, J.; Han, S.; Yu, J.; Park, Y.; Cho, S. H. *Angew. Chem. Int. Ed.***2023**, *62*, e202218794.
5. (a) Krautwald, S.; Sarlah, D.; Schafroth, M. A.; Carreira, E. M. *Science***2013**, *340*, 1065–1068. (b) Beletskaya, I. P.; Nájera, C.; Yus, M. *Chem. Rev.***2018**, *118*, 5080–5200. (c) Kim, U. B.; Jung, D. J.; Jeon, H. J.; Rathwell, K.; Lee, S.-G. *Chem. Rev.***2020**, *120*, 13382–13433.
6. Jung, Y.; Kim, J.; Kim, G.; Ko, J.; Hong, S.; Cho, S. H. *ACS Catal.***2024**, *14*, 18660–18668.

4D-STEM Analysis of Process-Dependent Cu-Sn Intermetallic Evolution in Micro-Bumps for Advanced Semiconductor Packaging

Sangjun Kang[1,2,3]

[1] In situ Electron Microscopy, Department of Materials Science, Technical University of Darmstadt, 64287 Darmstadt, Germany
[2] Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), 76344 Eggenstein-Leopoldshafen, Germany
[3] Karlsruhe Nano Micro Facility (KNMFi), Karlsruhe Institute of Technology (KIT), 76344 Eggenstein-Leopoldshafen, Germany
sangjun.kang@tu-darmstadt.de

Abstract. Cu/Sn-Ag micro-bumps are key metallic interconnects for high-density advanced semiconductor packaging, but joint reliability depends strongly on the morphology, thickness, and crystallographic structure of Cu-Sn intermetallic compounds (IMCs). Here, the process-dependent IMC microstructure produced by conventional reflow, pre-heating assisted reflow (H+Reflow), and laser-assisted soldering was compared using STEM-EDS and automated crystal orientation mapping based on 4D-STEM. The 4D-STEM phase maps directly identify a dual-layer reaction structure consisting of ε-Cu_3Sn adjacent to Cu and η-Cu_6Sn_5 toward the Sn-Ag solder, while orientation maps reveal pronounced differences in grain structure among the bonding routes. Reflow-based processing produces thicker and coarser IMC layers, with H+Reflow showing the most developed reaction structure because of its larger cumulative thermal budget. In contrast, laser-assisted soldering confines the interfacial reaction and preserves substantially finer IMC grains by limiting the time available for long-range interdiffusion and grain-boundary migration. STEM-EDS further shows process-dependent Ag redistribution near the solder/IMC region. These results demonstrate that bonding thermal history governs both the amount and crystallographic length scale of Cu-Sn IMCs and provide a microstructure-based guideline for advanced semiconductor packaging.

Keywords: advanced semiconductor packaging, Cu/Sn-Ag micro-bump, 4D-STEM, ACOM, Cu_3Sn, Cu_6Sn_5, reflow, laser-assisted soldering.

1. Introduction

Three-dimensional integration and heterogeneous packaging technologies are increasingly important for improving interconnect density, memory bandwidth, and system-level energy efficiency in advanced semiconductor devices [1,2]. Within fine-pitch micro-bumps, the interfacial microstructure occupies a large fraction of the joint volume and can therefore dominate electrical and mechanical reliability. Cu/Sn-based bonding is widely used because Cu is compatible with established metallization schemes and Sn-based solder provides low-temperature processability. During bonding, Cu and Sn react to form η-Cu_6Sn_5 and ε-Cu_3Sn [3-5]. These phases are required for metallurgical joining, but excessive growth is undesirable because Cu-Sn IMCs are more resistive and brittle than the parent metals [6,7].

The final IMC structure is strongly governed by the bonding thermal profile. Conventional reflow provides a relatively large thermal budget, pre-heating before reflow extends the effective reaction time, and laser-assisted soldering delivers highly localized energy over a short time scale [8-10]. Conventional SEM and EDS reveal morphology and elemental distribution, but they cannot fully resolve nanoscale phase identity, orientation, and grain structure. Automated crystal orientation mapping (ACOM) using 4D-STEM provides a phase- and orientation-resolved approach for this purpose [11,12]. In this study, Cu/Sn-Ag micro-bumps processed by reflow, H+Reflow, and laser-assisted soldering are compared to determine how thermal budget controls ε-Cu_3Sn/η-Cu_6Sn_5 phase distribution, thickness, and grain scale.

2. Experiments and Discussion

Cu/Sn-Ag micro-bumps were fabricated by electrochemical deposition of Cu and Sn-Ag solder. Conventional reflow was performed at 240 °C for 30 s. For H+Reflow, the specimen was first heated from 100 to 150 °C at 3 °C s^{-1} and then reflowed at 240 °C for 30 s. Laser-assisted soldering used a 980 nm laser pulse with a power

of 2 kW and a duration of 1 ms. Cross-sectional lamellae were prepared by focused ion beam milling and examined by HAADF-STEM and STEM-EDS. ACOM 4D-STEM datasets were acquired at 300 kV with a DECTRIS Quadro detector using a 0.5 mrad convergence semi-angle, 460 mm camera length, and 1 ms acquisition per probe position. Diffraction patterns were indexed in ASTAR against templates for Cu, β-Sn, ε-Cu_3Sn (Pmmn), η-Cu_6Sn_5 (C2/c), and Ag_3Sn.

The 4D-STEM maps in Fig. 1 directly resolve the process-dependent crystallographic structure of the interfacial reaction layer. In every bonded specimen, ε-Cu_3Sn is located next to the Cu side and η-Cu_6Sn_5 toward the Sn-Ag solder, confirming a dual-IMC architecture. Reflow and H+Reflow produce a thicker and more laterally continuous reaction region, while laser-assisted soldering confines both IMC phases to a narrower interface. The orientation maps also show that reflow-based processing promotes larger crystallographic domains, whereas the laser condition preserves a more refined and fragmented IMC grain structure. Complementary STEM-EDS supports the phase sequence through the Cu/Sn concentration gradient and shows local Ag redistribution within the Sn-rich region and near selected interfacial features.

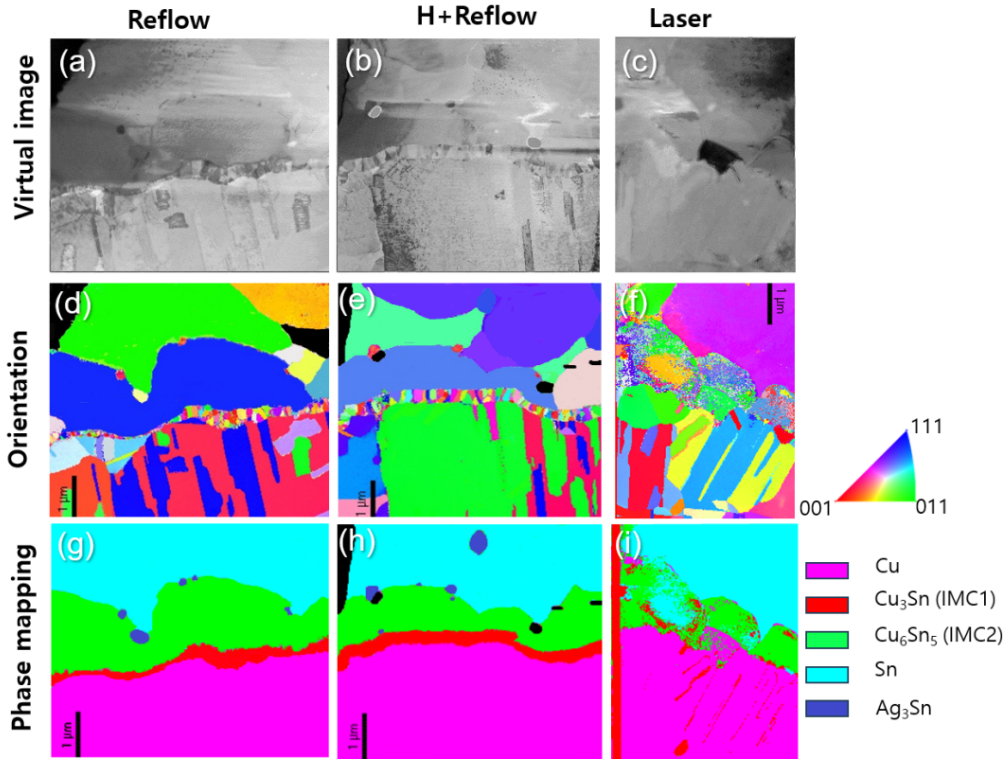


Fig. 1. ACOM 4D-STEM analysis of Cu/Sn-Ag micro-bumps after reflow, H+Reflow, and laser-assisted soldering. Virtual images (a-c), crystallographic orientation maps (d-f), and phase maps (g-i) identify Cu, ε-Cu_3Sn, η-Cu_6Sn_5, Sn, and Ag-containing regions. Laser-assisted soldering produces a more confined and refined IMC structure.

Quantitative analysis (Fig. 2) shows that IMC thickness and grain size follow the same process dependence. Both ε-Cu_3Sn and η-Cu_6Sn_5 are substantially thicker after reflow-based processing than after laser soldering, and the H+Reflow condition gives the largest reaction layers. The grain-size analysis similarly shows pronounced coarsening after reflow and H+Reflow, while the laser condition yields markedly finer grains. This correlation demonstrates that reaction-front migration and grain coarsening are coupled consequences of the effective thermal budget. Pre-heating extends the period available for solid-state interdiffusion before the main reflow step, whereas the short laser pulse limits long-range diffusion and grain-boundary migration.

From a reliability perspective, the desirable joint contains a continuous IMC layer sufficient for metallurgical bonding while avoiding excessive volumes of brittle and resistive Cu-Sn reaction products. Reflow and H+Reflow provide robust interfacial reaction but promote IMC thickening and coarsening. Laser-assisted soldering suppresses these kinetic processes and therefore acts as a distinct microstructure-control strategy, although its processing window must still ensure complete wetting and a continuous bonded interface. The combined STEM-EDS and 4D-STEM analysis provides a crystallography-resolved framework for evaluating this trade-off.

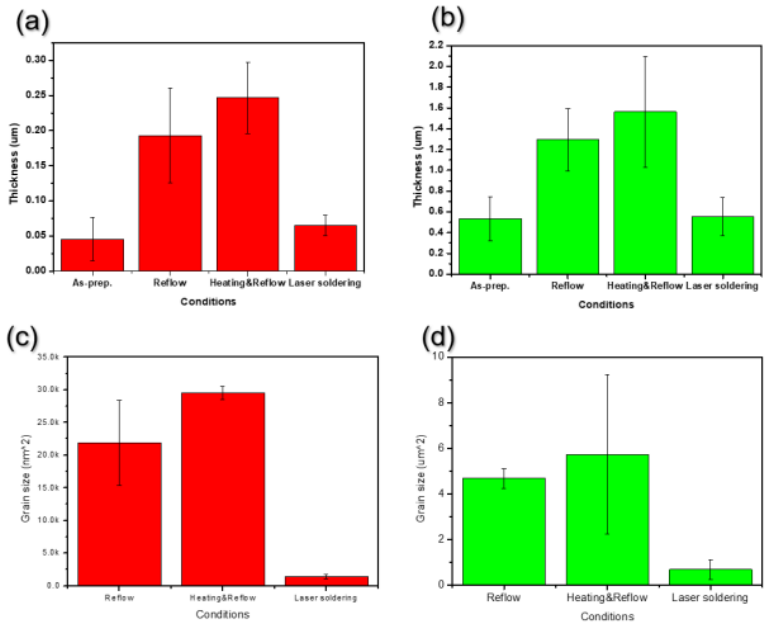


Fig. 2. Process-dependent thickness of (a) ε-Cu_3Sn and (b) η-Cu_6Sn_5, and grain size of (c) ε-Cu_3Sn and (d) η-Cu_6Sn_5. Reflow-based processing gives thicker and coarser IMCs, while laser-assisted soldering suppresses both growth and coarsening.

3. Conclusions

ACOM 4D-STEM shows that bonding thermal history controls both phase distribution and grain scale in Cu/Sn-Ag micro-bumps. Reflow and H+Reflow produce thicker and coarser ε-Cu_3Sn/η-Cu_6Sn_5 layers, whereas localized ultrafast laser heating confines the reaction and preserves a finer IMC structure. These crystallography-resolved results support thermal-budget optimization of micro-bumps for advanced semiconductor packaging.

References

1. Das Sharma, D., Pasdast, G., Tiagaraj, S., Aygun, K.: High-performance, power-efficient three-dimensional system-in-package designs with universal chiplet interconnect express. Nature Electronics 7, 244–254 (2024).
2. Lau, J.H.: Recent advances and trends in advanced packaging. IEEE Trans. Components, Packaging and Manufacturing Technology 12, 228–252 (2022).
3. Annuar, S., Mahmoodian, R., Hamdi, M., Tu, K.N.: Intermetallic compounds in 3D integrated circuits technology: a brief review. Science and Technology of Advanced Materials 18, 693–703 (2017).
4. Liu, H., Wang, K., Aasmundtveit, K., Hoivik, N.: Intermetallic compound formation mechanisms for Cu-Sn solid-liquid interdiffusion bonding. J. Electronic Materials 41, 2453–2462 (2012).
5. Kannojia, H.K., Dixit, P.: A review of intermetallic compound growth and void formation in electrodeposited Cu-Sn layers for microsystems packaging. J. Materials Science: Materials in Electronics 32, 6742–6777 (2021).
6. Frederikse, H.P.R., Fields, R.J., Feldman, A.: Thermal and electrical properties of copper-tin and nickel-tin intermetallics. J. Applied Physics 72, 2879–2882 (1992).
7. Balakrisnan, B., Chum, C.C., Li, M., Chen, Z., Cahyadi, T.: Fracture toughness of Cu-Sn intermetallic thin films. J. Electronic Materials 32, 166–171 (2003).
8. Dong, C., Shang, M., Chen, F., Wang, Y., Li, X., Ma, H., Ma, H.: Investigation on growth of the orientation-preferred Cu_6Sn_5 on (001) Cu during the temperature-increased multiple reflow process. J. Alloys and Compounds 885, 161205 (2021).
9. Tatsumi, H., Kaneshita, S., Kida, Y., Sato, Y., Tsukamoto, M., Nishikawa, H.: Highly efficient soldering of Sn-Ag-Cu solder joints using blue laser. J. Manufacturing Processes 82, 700–707 (2022).
10. Nishikawa, H., Iwata, N.: Improvement of joint reliability of Sn-Ag-Cu solder bumps on Cu by a laser process. Materials Transactions 56, 1025–1029 (2015).
11. Rauch, E.F., Véron, M.: Automated crystal orientation and phase mapping in TEM. Materials Characterization 98, 1–9 (2014).
12. Savitzky, B.H., et al.: py4DSTEM: a software package for four-dimensional scanning transmission electron microscopy data analysis. Microscopy and Microanalysis 27, 712–743 (2021).

Advanced Synchrotron X-ray Scattering Facility and Operando Materials Dynamics at the PLS-II 3C SAXS-I Beamline

Young Yong Kim[1]

[1] Pohang Accelerator Laboratory (PAL), POSTECH, Pohang, Gyeongbuk 37673, Republic of Korea
kimyy@postech.ac.kr

Abstract. This paper presents a comprehensive overview of the analytical capabilities, specialized infrastructure, and recent operando scientific breakthroughs hosted at the PLS-II 3C SAXS-I beamline of the Pohang Accelerator Laboratory (PAL). Powered by an in-vacuum undulator insertion device and a high-resolution double-crystal monochromator, the 3C SAXS-I beamline delivers high-brilliance hard X-rays (8.3 - 20.6 keV) optimized for multi-dimensional structural characterization ranging from tens of nanometers down to sub-nanometer atomic dimensions(~3Å). The beamline features an advanced experimental infrastructure, including a machine-learning (ML)-driven, grazing-incidence (GI) automated sample-alignment system, a custom in situ electrical biasing setup, an ultra-fast Eiger X4M single-photon-counting 2D detector, and a customized data-reduction processor. Representative operando research highlights are detailed: (1) reversible electrochemical ion-intercalation mechanics in aqueous zinc-ion battery cathodes (Na2V6O16 · 2H2O), (2) pulse-driven transient microstructural kinetics in organic mixed ionic-electronic conductors (OMIECs) via time-aligned operando GIWAXS, and (3) moisture-induced reversible dynamic lattice rearrangement in ethylammonium-based perovskites. Finally, a roadmap toward an intelligent multi-stage beamline control and single-step data analysis platform is presented.

Keywords: Synchrotron radiation, Small-Angle X-ray Scattering (SAXS), Wide-Angle X-ray Scattering (WAXS), Operando characterization, Machine learning alignment, Battery cathodes, Organic mixed ionic-electronic conductors, Perovskite phase transition.

1. Introduction

In modern advanced materials science and engineering, the primary analytical paradigm has decisively shifted toward real-time observation in operational environments (in situ and operando characterization). Dissecting non-equilibrium dynamic mechanisms—such as structural phase transitions in secondary battery electrodes, transient ion transport in mixed ionic-electronic conductors for neuro-inspired computing, and moisture-induced lattice rearrangements in halide perovskite optoelectronics—requires highly specialized synchrotron X-ray scattering platforms capable of capturing rapid structural dynamics across multiple spatial and temporal scales.

To meet these demanding analytical requirements, the PLS-II 3C SAXS-I beamline at the Pohang Accelerator Laboratory (PAL) provides an advanced hard X-ray scattering environment. By integrating an in-vacuum undulator insertion device with precision monochromatization, the beamline delivers stable, high-brilliance X-ray beams across the hard X-ray regime. This article presents a comprehensive technical description of the beamline instrumentation, intelligent automation software, custom experimental sample environments, and key operando scientific milestones in next-generation energy, electronic, and optoelectronic materials.

2. Beamline Instrumentation and Advanced Infrastructure

The 3C SAXS-I beamline delivers high-flux hard X-rays (~1012 photons/sec) with tunable photon energies from 8.3 to 20.6 keV (routinely operated at baseline energies of 8.95 keV and 11.27 keV). The facility accommodates diverse non-solution scattering geometries across solid-state and organic thin-film systems:

- **Small-Angle X-ray Scattering (SAXS)**: Characterizes long-range periodicities, macromolecular self-assembly, nanoparticle morphology, and domain structures (1-100 nm).
- **Wide-Angle X-ray Scattering (WAXS)**: Probes sub-nanometer crystal lattice constants, crystallite orientation, and phase evolution down to atomic dimensions (~3 Å).

- **Grazing-Incidence SAXS/WAXS (GISAXS/GIWAXS)**: Surface- and interface-sensitive scattering geometries dedicated to ultra-thin organic semiconductors, display functional layers, and perovskite active layers (< 200 nm thickness).

2.1 Specialized Hardware and Software Infrastructure

- **Machine-Learning (ML)-Driven Automated GI Sample Alignment System**: Grazing-incidence scattering requires sub-millidegree alignment precision. A reinforcement-learning-based control algorithm was deployed to achieve fully automated, rapid sample alignment, eliminating human operational errors and enhancing effective beamtime throughput.
- **Custom In-Situ Operando Electrical Biasing Apparatus**: A specialized environmental chamber capable of applying synchronized electrical pulses during GIWAXS measurements was developed. This setup allows continuous monitoring of dynamic ion kinetics and structural rearrangements in functional electronic thin films under active bias.
- **Eiger X4M(Dectris) Single-Photon Counting Detector Integration**: Featuring sub-millisecond frame rates and extreme dynamic range, the hybrid pixel 2D detector captures rapid lattice strain evolution and transient intermediate phases with minimal noise floor.
- **Customized One-Stop Data Reduction Pipeline**: A dedicated software suite converts large-volume 2D diffraction image sequences into azimuthally integrated 1D scattering profiles (I(q) vs. q) in real time, supporting custom geometrical corrections and automated profile fitting.

3. Key Operando Research Achievements

3.1 Operando Reversible Structural Dynamics in Aqueous Zinc-Ion Battery Cathodes

Aqueous zinc-ion batteries (AZIBs) represent a promising alternative for large-scale energy storage due to low cost and high safety. Utilizing real-time transmission-mode operando WAXS, the structural intercalation dynamics of Na2V6O16 · 2H2O (NaVO) cathodes prepared via sonochemical synthesis were tracked during electrochemical charge/discharge cycles [1].

The operando diffraction maps revealed flexible expansion and contraction of the c-axis lattice parameter corresponding to Zn2+ insertion and extraction. The structural integrity was preserved across extended cycling due to the interlayer Na+ ions acting as robust "structural pillars". This study directly validated the origin of high cycling reversibility, as published in the Journal of Materials Chemistry A (2025) [1].

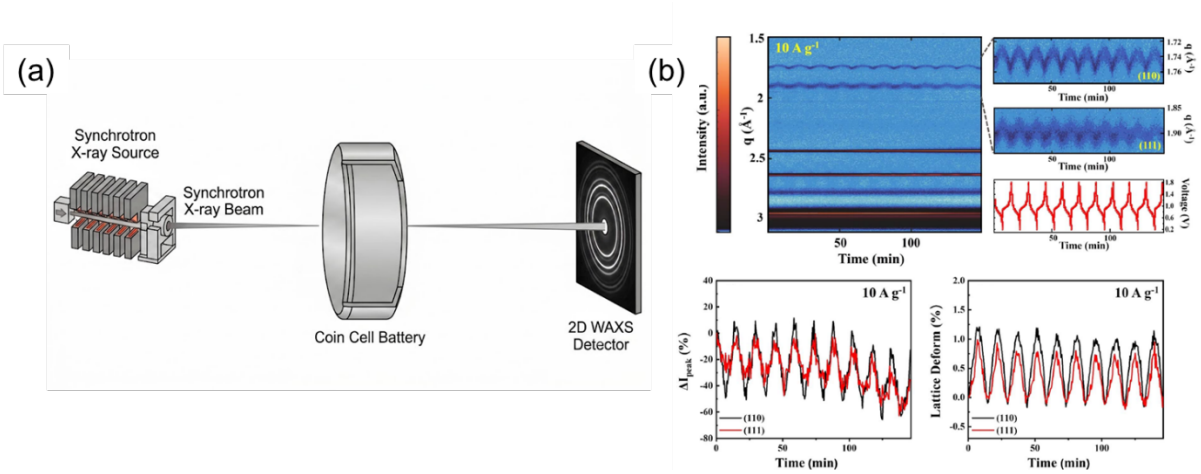


Fig. 1. Operando structural analysis of an aqueous zinc-ion battery (AZIB) employing a Na2V6O16 · 2H2O (NaVO) cathode performed at the PLS-II 3C SAXS-I beamline. (a) Schematic illustration of the transmission-mode synchrotron X-ray scattering setup, displaying the alignment of the high-energy X-ray beam, the vertical battery coin cell, and the 2D WAXS detector. (b) Real-time contour map tracking the reversible shift of scattering vector q for the (110) and (111) diffraction peaks (q ≈ 1.75 Å-1 and 1.94 Å-1) synchronized with galvanostatic charge/discharge voltage profiles.

3.2 Pulse-Driven Transient Microstructural Kinetics in Organic Mixed Ionic-Electronic Conductors

Organic mixed ionic-electronic conductors (OMIECs) are pivotal components in neuromorphic computing devices. However, capturing transient structural reorganizations under brief voltage pulses is challenging due to detector acquisition speed limits relative to fast ionic transport.

To resolve this temporal disparity, a novel time-aligned operando GIWAXS technique was established [2]. By engineering a custom long-channel device architecture (d = 4.5 mm gate-to-channel distance), the effective ion diffusion time constant was deliberately expanded (τ ~ 50 - 75 s). Under voltage pulse excitation, real-time GIWAXS captured the transient influx of TFSI- anions into the lamellar domains, revealing a previously undetected metastable intermediate state pathway during non-equilibrium doping [2].

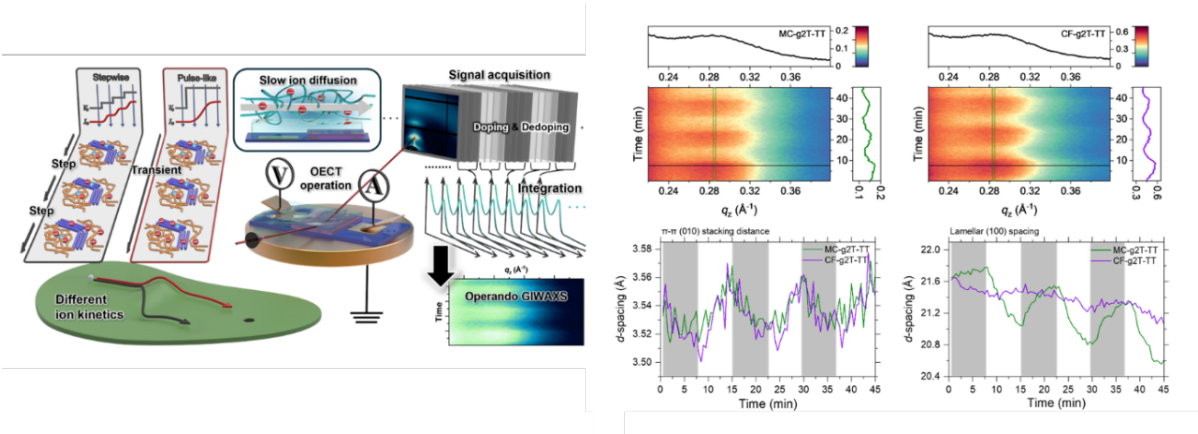


Fig. 2. Pulse-driven anion penetration dynamics in OMIEC thin films investigated via high-speed operando GIWAXS established at the PLS-II 3C SAXS-I beamline. Schematic depiction of the time-aligned operando GIWAXS measurement strategy and transient ion transport kinetics in the organic electrochemical transistor (OECT) channel (left), accompanied by time-resolved 2D intensity contour maps of q_z and real-time evolution of lamellar (100) and π- π (010) *d*-spacings under pulsed voltage modulation (right).

3.3 Reversible Dynamic Lattice Rearrangement in Ethylammonium-Based Perovskites

The moisture-driven phase behavior of ethylammonium (EA)-based perovskite thin films ([EAI]/[PbI2] ≥ 1.6) was systematically investigated via in-situ GIWAXS [3]. Contrary to the typical irreversible degradation into PbI2 observed in conventional methylammonium perovskites, EA-based perovskites exhibit a topotactic, degradation-free transition between 1D chain structures and 2D layered structures upon humidity exposure.

The continuous diffraction tracking demonstrated rapid recovery of the original crystal frame upon environmental cycling. This "dynamic lattice rearrangement" provides foundational principles for smart climate-responsive windows and permanent moisture history sensors [3].

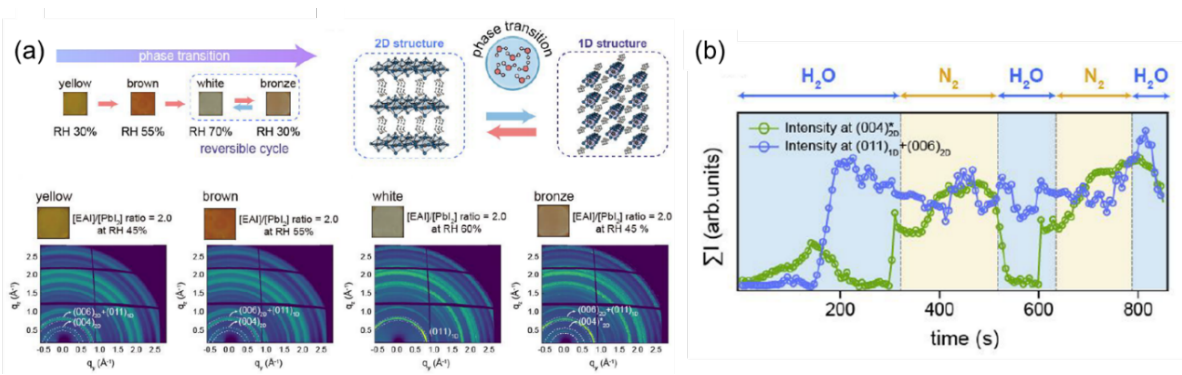


Fig. 3. Real-time in-situ observation of moisture-responsive dynamic lattice rearrangement in ethylammonium (EA) perovskites governed by stoichiometry control. (a) Four-step color-transition model of reversible multichromism and schematic mechanism of the dynamic 2D-to-1D lattice transformation under controlled relative humidity (RH). (b) Time-resolved diffraction peak intensity profiles displaying the complete suppression of the 2D-associated (004)*2D reflection and the simultaneous predominance of the 1D crystalline chain (011)1D peak during cyclic H2O and N2 exposure.

4. Beamline Technical Specifications

Parameter	Specification Details
X-ray Source	PLS-II In-Vacuum Undulator (1.4m short, 20 mm period)
Monochromator	Double Crystal Si(111) System
Photon Energy Range	8.3-20.6 keV (Routinely operated at 8.95keV and 11.27 keV)
Photon Flux	~10^{12} photons/sec
Beam Size at Sample	60x350μm^2 (VerticalμHorizontal)
Sample-to-detector distance	0.2 - 5.0 m
2D Detector System	Dectris Eiger X 4M Single-Photon Counting Detector
Supported Scattering Modes	SAXS, WAXS, GISAXS, GIWAXS (Solid-state & thin-film forms)
Specialized In-Situ Infrastructure	ML-driven automated sample alignment, operando electrical biasing apparatus

5. Conclusions and Future Development Roadmap

The PLS-II 3C SAXS-I beamline serves as a vital user platform supporting advanced materials research across energy storage, neuromorphic electronics, and functional perovskites. Future developments target the implementation of a Multi-Stage Modular Docking System to enable rapid interchange of complex sample environments. Furthermore, an AI-assisted beamline control interface and an integrated smart automated curve-fitting suite are under active development to streamline data processing from beamtime proposal to final quantitative analysis.

References

1. E. Lee, J. Lee, Y. So, H. Seo, S. H. Lee, J. Kang, Y. Y. Kim, B. H. Kim, and S. Mhin, “Enhanced electrochemical performance of aqueous Zn-ion batteries based on Na2V6O16 · 2H2O cathodes: insights from DFT and synchrotron X-ray analysis”, Journal of Materials Chemistry A, vol. 13, no. 12, pp. 6789–6801, 2025.
2. T. W. Yoon, J. Lee, J. Noh, B. Kim, B. Kim, S. Park, Y. Y. Kim, and B. Kang, “Time-Aligned Operando GIWAXS for Capturing Pulse-Driven Structural Dynamics in Organic Mixed Ionic Electronic Conductors”, InfoMat (Under Review, 2026).
3. A. Ha, S. Kim, K. Kim, E. Yoon, S. Pak, Y. Yi, D. Shin, Y. Y. Kim, K. H. Lee, and S. Park, “Turning instability into functionality: Stoichiometry- and moisture-tuned reversible multichromism in Ethylammonium perovskites”, Nature Communications (Under Review, 2026; preprint on Research Square, DOI: 10.21203/rs.3.rs-9305944/v1).

Study of Electrochemistry via In-situ Transmission Electron Microscopy

Junbeom Park[1]

[1]Institute of Energy Technologies - Fundamental Electrochemistry (IET-1),
Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52425 Jülich, Germany
j.park@fz-juelich.de

Abstract. With the accelerating development of batteries and green hydrogen technologies, electrochemical research has expanded rapidly. However, conventional macro-scale characterization at the cell or stack level faces clear limitations in further driving device efficiency. Since fundamental electrochemical processes are governed by interactions among electrons and ions at the nanoscale, unveiling these nanoscale mechanisms is crucial for performance enhancements. Among various nanoscale analysis tools, transmission electron microscopy (TEM) is uniquely capable of multi-modal analysis, combining high-resolution imaging with chemical characterization. Notably, recent advancements in in-situ TEM enable real-time tracking of dynamic electrochemical processes directly within the microscope, providing deeper insights beyond static ex-situ comparisons. In this review, we highlight key methodologies in in-situ liquid-biasing TEM and subsequent image processing workflows for quantitative analysis of dynamic electrochemical phenomena. By bridging nanoscale observations with quantitative metrics, this work shows the role of in-situ TEM in advancing future electrochemical energy storage and conversion devices.

Keywords: In-situ transmission electron microscopy, Image processing, Electrochemistry

1. Introduction

The urgent global demand for scalable clean energy storage and conversion devices, such as batteries and green hydrogen, has increased an unprecedented volume of electrochemical research. Despite significant macroscopic advancements, conventional diagnostic methods operating at the full-cell or stack scale has been showing low-efficiency to enhancing the performance. Macro-scale measurements, such as global current-voltage metrics or bulk impedance spectroscopy, integrate all the local phenomena and often ignore the spatial and temporal heterogeneities that govern degradation and energy efficiency.

Fundamental electrochemical processes, including phase transformations, mass transport, interface passivation, and electrodeposition, are inherently governed by local ion-electron interactions occurring at the nanoscale. To overcome the limitations of macroscopic characterization and achieve rational design of high-performance energy devices, it is essential to investigate these localized dynamic mechanisms in real time.

Among various advanced analytical tools, transmission electron microscopy (TEM) stands out due to its unique characterization abilities such as high-resolution imaging and elemental mapping. Furthermore, the advent of *in-situ* and *operando* liquid-phase TEM (LPTEM) using micro-electro-mechanical systems (MEMS)-based liquid cells has revolutionized the field. Instead of assuming reaction mechanisms through static *ex-situ* post-mortem snapshots, *in-situ* liquid-biasing TEM allows researchers to track full dynamic electrochemical cycles directly inside the microscope under applied electrical bias.

In this review, we examine the principles and recent progress in *in-situ* liquid-biasing TEM, demonstrating how this technique transforms qualitative nanoscale observations into rigorous quantitative evaluations. By featuring representative workflows based on high-angle annular dark-field (HAADF) scanning TEM (STEM) and targeted image processing, we highlight how *in-situ* methodology bridges nanoscale findings with macro-scale device performance.

2. Experiments and Discussion

To conduct electrochemical experiments inside the TEM without compromising the ultra-high vacuum environment, MEMS-based liquid cell holders equipped with electron-transparent silicon nitride (Si_3N_4) windows are utilized. Microelectrodes with Platinum (Pt) are integrated into the MEMS chip, enabling precise electrical biasing during imaging. Liquid electrolyte flows to Pt electrode through liquid cell holder, but it is isolated to ultra-high vacuum environment.

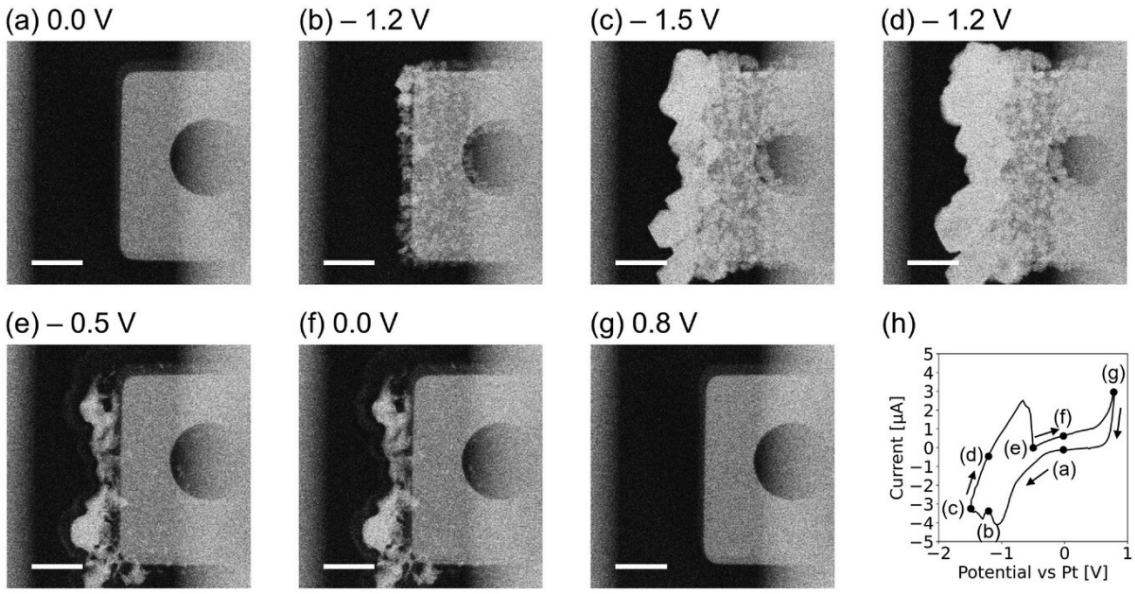


Fig. 1. a–g) HAADF STEM images during Zn plating (a–d, at 0.0, −1.2, −1.5, and −1.2 V) and stripping (e–g, at −0.5, 0.0, and 0.8 V) process in 0.1 m $ZnSO_4$ solution on the finger-shaped Pt working electrode during the cyclic voltammetry. The scale bars in (a–g) are 2 μm. h) The corresponding Current versus Potential curve during CV. The black dots on the CV curve in (h) indicate the biasing condition at the corresponding STEM images shown in (a–g). The flow rate of the 0.1 M $ZnSO_4$ was 1.7 μL min^{-1}, and the potential scan rate was 0.1 V s^{-1} [1]

Fig. 1 illustrates a representative *in-situ* liquid-biasing TEM study using HAADF-STEM imaging during a cyclic voltammetry (CV) sequence. Because HAADF-STEM intensity scales with atomic number (*Z*-contrast), the dynamic formation and dissolution of metallic deposits (bright white features) on the working electrode (rectangle with round hole) can be clearly distinguished from the surrounding liquid electrolyte and substrate (black background). As the electrical bias crosses the reduction potential, initial nucleation sites appear dynamically along the electrode surface. During the reverse anodic sweep, the continuous stripping process and morphological dissolution are captured in real-time frame-by-frame series.

While qualitative visual tracking offers crucial direct insights, extracting quantitative metricssuch as growth ratesis vital for establishing predictive electrochemical models.Fig.2 demonstrates an advanced quantitative workflow applied to *in-situ* TEM movie frames recording dendritic growth. By applying basic image processing techniquesincluding noise reduction, background subtraction, and binary segmentation, the precise area and relative volume of deposited metal can be isolated frame-by-frame.

As depicted in Fig.2(e–g), second-by-second frame subtraction creates spatial growth maps that visually color-code the incremental expansion of individual dendrites. This segmentation allows to obtain1) linear growth rate, 2) volumetric deposition rate and 3) symmetry and tip kinetics. This quantitative methodology transforms normal image series into selective electrodeposition kinetics, enabling direct comparisons with macroscopic electrochemical current efficiencies.

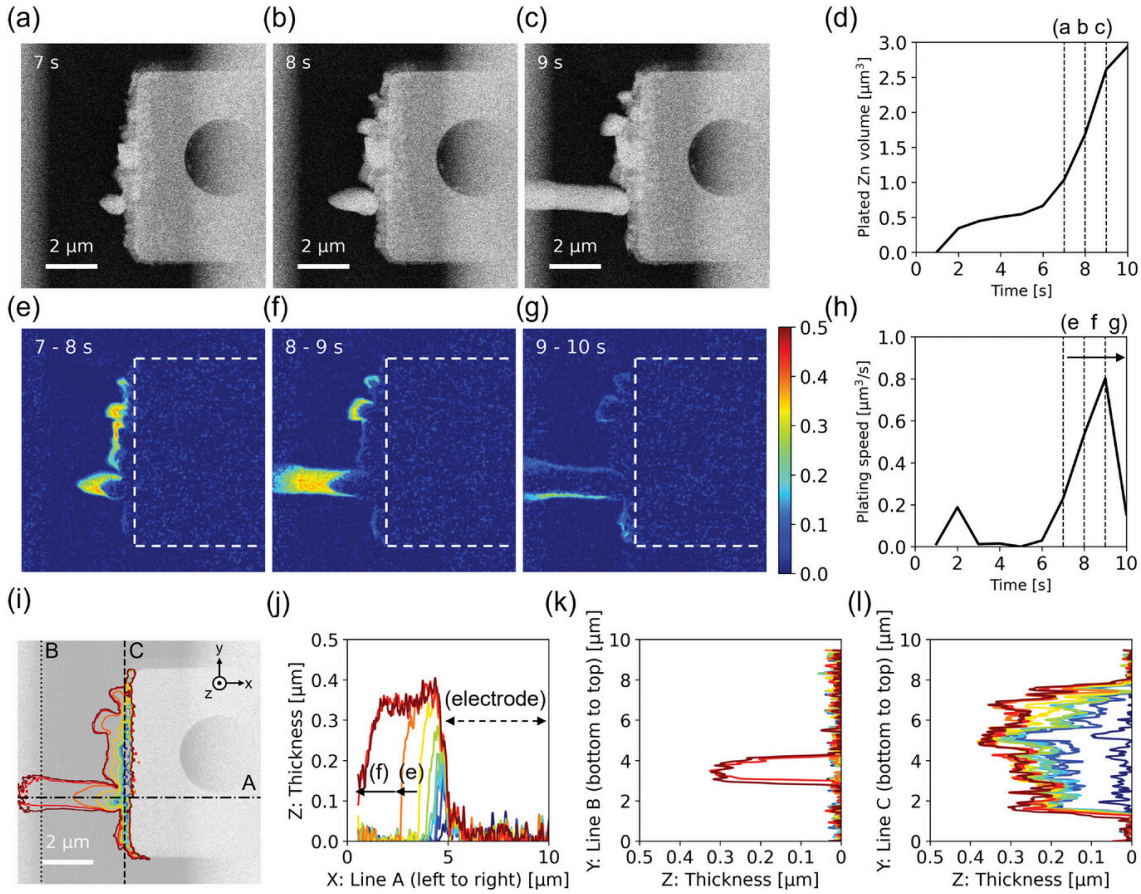


Fig. 2. Investigation of Zn dendritic growth during plating in static 0.1 m $ZnSO_4$. a–c) STEM images obtained at seventh, eighth, and ninth seconds. d) The plated Zn volume versus time. e–g) Processed images show the Zn dendritic growth between 7–8, 8–9, and 9–10 s. A color map on (g) represents deposit thickness at eighth (e), ninth (f), and 10th (g) second. The white dotted rectangle in (e–g) represents the Pt electrode. h) The plating speed of Zn versus time. i–l) Tracking profiles of Zn dendritic growth: i) along X-Y plane, top-view of the dendrite, j) along X-Z plane, thickness versus horizontal dash-dot line A in (i), k) along Y-Z plane, thickness versus vertical dotted line B in (i), and (l) along Y-Z plane, thickness versus vertical dashed line C in (i). The profile's color indicates the time from 1s (blue) to 10 s (red). The background image in (i) is pristine Pt electrode [1]

3. Conclusions

In-situ liquid-biasing TEM has evolved from a qualitative visualization tool into a quantitative diagnostic discipline for modern electrochemistry. By providing real-time observations of ion-electron interactions at sub-nanometer spatial scales, this technique fills the critical knowledge gap between fundamental nanoscale physics and macroscopic energy storage/conversion performance.

Through integrated image processing workflows—such as background subtraction and threshold segmentation applied to time-resolved HAADF-STEM data—key electrodeposition metrics, including volume plating rate and dendrite growth velocity, can be precisely quantified. Combining these *in-situ* liquid-biasing setups with advanced analytical techniques like 4D-STEM will further expand our understanding of crystallographic orientations, local strain, and interfacial chemistry. Ultimately, the quantitative methodologies established through *in-situ* TEM will play a pivotal role in guiding the rational design of next-generation battery anodes and durable fuel cell/electrolyzer catalysts.

References

1. J. Park et al., Small methods 8 (2024) 2400081 DOI:10.1002/smtd.202400081
2. J. Park et al., Advanced Materials Interfaces 26 (2024) 2302146 DOI: 10.1002/adem.202302146
3. P. Chakraborty et al., Small methods (2026) DOI: 10.1002/smtd.70809

Powering the world of microrobots with micro energy systems

Yeji Lee[1,2], Vineeth Kumar Bandari[1,2]and Oliver G. Schmidt[1,2]

[1] Material Systems for Nanoelectronics, Chemnitz University of Technology, 09107 Chemnitz, Germany.
[2] Research Center for Materials, Architectures and Integration of Nanomembranes (MAIN), Chemnitz University of Technology, 09126 Chemnitz, Germany.
yeji.lee@main.tu-chemnitz.de

Abstract. The growing demand for miniaturized energy sources for microelectronic and biomedical applications has accelerated the development of compact, flexible, and high-performance micro-energy systems. This work explores the integration of Swiss-roll self-assembly and origami-inspired fabrication to realize multifunctional microsystems with integrated energy harvesting, storage, sensing, communication, and actuation capabilities.

Keywords: Microelectronics, Micro energy storage/harvesting system, Nano-biosupercapacitor, Micro-organic solar cell, Modular micro-origami robot

1. Introduction

The advancements in microrobotics have progressed significantly from Feynman's visionary concepts [1], leading to the development of self-propelled catalytic microengines [2] and advanced MEMS-based microsystems. A notable example is the motile twin-jet-engine microsystem (MTJEMS) [3], which integrates rolled-up catalytic engines with wireless energy transfer to achieve stable and autonomous locomotion. However, realizing fully autonomous microrobots remains challenging, as it requires the seamless integration of locomotion, sensing, computation, actuation, and onboard energy supply within a compact and flexible architecture.

Among these functionalities, microscale energy storage and harvesting are particularly critical for enabling untethered operation. The growing demand for miniaturized energy sources in microelectronics and biomedical applications has therefore driven the development of efficient, scalable, and flexible energy technologies. In this context, Swiss-roll self-assembly and origami-inspired fabrication have emerged as promising strategies for creating compact, mechanically flexible, and high-performance energy systems.

These technologies provide the foundation for modular micro-origami robots, or "smartlets," equipped with diverse functionalities, energy harvesters, sensors, microprocessors, micro-LEDs, and actuators into millimeter-scale autonomous platforms [4].

2. Experiments and Discussion

The realization of autonomous microrobots requires the integration of multiple functionalities, including energy generation, sensing, computation, communication, and actuation, within a compact and scalable architecture. However, conventional microsystems are often limited by restricted surface area, which hinders the integration of diverse components and the development of fully autonomous platforms.

Swiss-roll self-assembly and origami-inspired fabrication have emerged as promising approaches for creating compact, mechanically flexible, and high-performance microsystems. Nevertheless, efficient energy management remains a critical challenge for autonomous smartlets, as reliable energy storage is essential to maintain operation under varying environmental conditions and to complement energy harvesting systems.

To address this challenge, a fully biocompatible nano-biosupercapacitor (nBSC) is developed using Swiss-roll self-assembly [5], as shown in **Fig 1**. The tubular architecture enables an ultracompact energy storage device with a volume of only 1 nL while maintaining functionality in biological environments. The nBSC stores energy in physiological electrolytes, including blood, and delivers up to 1.6 V. Its tubular geometry provides mechanical protection against external stresses, such as fluid flow and deformation, while bioelectrochemical interactions enhance performance and mitigate self-discharge.

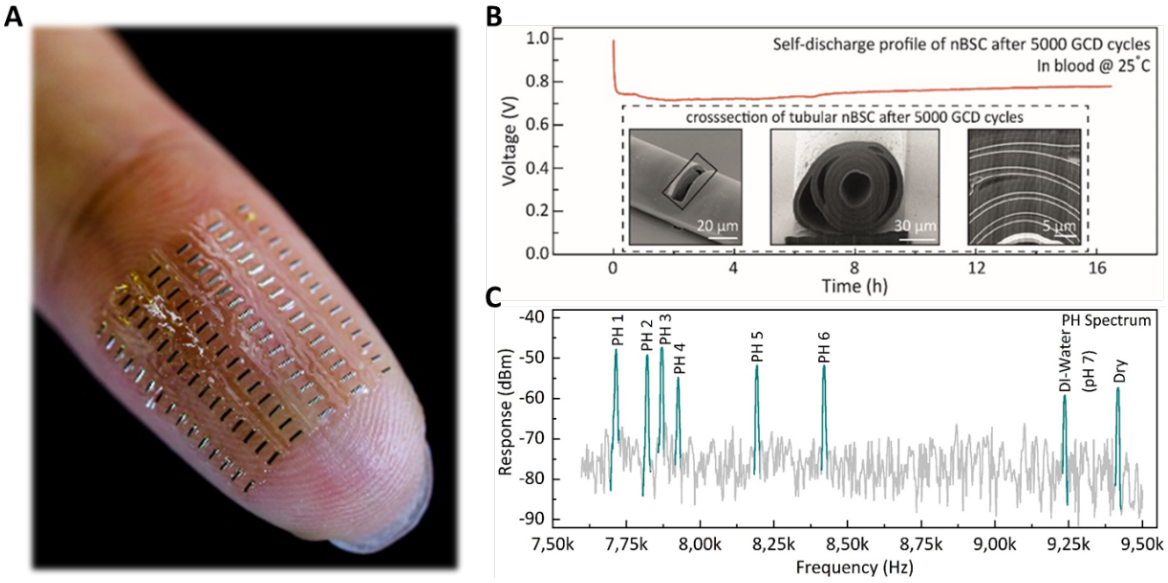


Fig. 1. Nano-biosupercapacitors (nBSCs) enable autarkic sensor operation in blood. **A,** Photograph of a 6 x 15 nBSCs array on a fingertip. **B,** Self-discharge profile in blood after 5000x GCD cycles. Inset shows cross-sectional images of tubular nBSC after 5000x GCD cycles. (All measurements are performed under ambient condition at 25 °C.). **C,** Frequency spectral response of the nBSC based pH sensor at various electrolyte pH.

The incorporation of biological components further improves the device functionality, with redox-active enzymes and living cells naturally present in blood enhancing electrochemical performance by approximately 40%. The nBSC demonstrates stability under physiological conditions and excellent biocompatibility, enabling its integration into biomedical microsystems. Furthermore, the stored energy is sufficient to operate an integrated pH sensor, demonstrating its potential for powering functional electronics in confined biological environments.

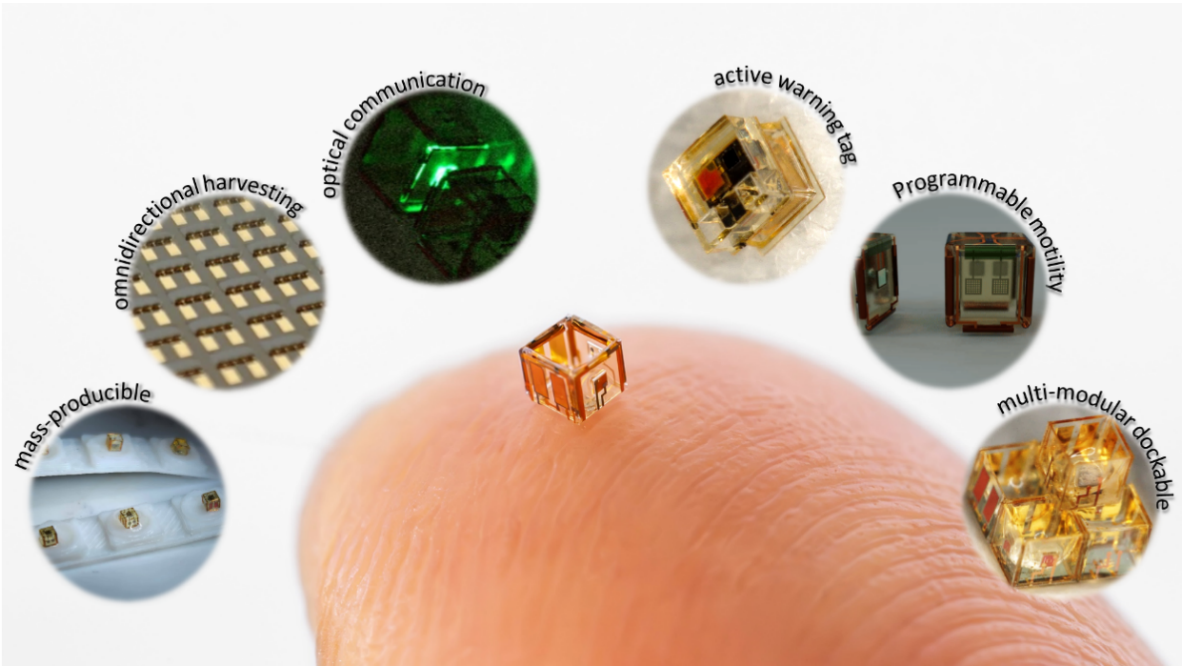


Fig. 2. Smartlet integrating multiple functionalities; mass-producible modular design, omnidirectional solar energy harvesting, optical communication, active warning signaling, programmable locomotion, and multimodular docking for collective operation, in a 1 mm platform.

The smartlet architecture integrates several essential robotic functionalities within a single platform [4], as shown in **Fig 2**. Micro-organic solar cells (μOSCs) positioned along the edges of the structure provide omnidirectional energy harvesting, enabling operation under ambient light conditions. The rolled-up geometry increases the effective light absorption area and reduces the dependence on incident light direction compared with conventional planar devices. Micro-organic photodetectors (μOPDs) are integrated as optical sensors for orientation detection, while silicon chiplets provide onboard computation and control. In addition, micro-LEDs enable optical communication between individual modules, creating a pathway toward distributed and cooperative microrobotic systems.

A major achievement of the smartlet platform is the successful integration of rigid microelectronic components onto flexible and foldable structures. The incorporation of silicon chiplets requires reliable electrical interconnection across deformable polymer substrates while preserving mechanical flexibility. Through heterogeneous integration and microchip bonding approaches, the smartlets achieve compact electronic functionality without compromising the self-folding process. This demonstrates that conventional semiconductor technologies can be combined with soft, three-dimensional architectures to create highly integrated microsystems.

The developed smartlets further demonstrate autonomous operation through integrated actuation mechanisms. Electrochemical bubble-generating electrodes provide controlled buoyancy changes, allowing programmed locomotion in aqueous environments. The onboard processor controls these actuators based on sensor inputs and programmed instructions, demonstrating the transition from passive microstructures toward active robotic systems. Furthermore, optical communication through integrated micro-LEDs enables information exchange between modules, supporting future collective behaviors similar to biological systems.

3. Conclusions

This work advances autonomous microsystems through the development and integration of miniaturized energy storage, energy harvesting, and multifunctional origami-based platforms. This research presents a fully biocompatible nano-biosupercapacitor, designed to operate in biological environments. Featuring a self-assembled Swiss-roll architecture, it occupies just 1 nanoliter of volume and delivers up to 1.6 V in blood. The tubular geometry not only imparts mechanical robustness but also leverages redox-active enzymes and blood cells to enhance performance by 40%, while significantly reducing self-discharge. When fully charged, it successfully powers a pH sensor system in situ. Furthermore, a micro-organic solar cell based on a P3HT:$PC_{61}BM$ blend in a Swiss-roll configuration achieves 7% power conversion efficiency and exhibits excellent angular light absorption, addressing key limitations in organic photovoltaics such as angle dependence and long-term stability, making it suitable for powering micro-actuators.

Finally, the work demonstrates the self-assembly of strain-engineered thin membranes into 3D origami-cube smartlets. These miniature modules integrate functionalities such as solar energy harvesting, electrolysis-based locomotion, sun-oriented photodetection, and optical inter-module communication. At just 1 mm in size, these multifunctional smartlets showcase the transformative potential of autonomous, modular microrobotic systems for future applications in medicine, sensing, and beyond.

References

1. Feynman, R.P. There´s plenty of room at the bottom. J. Microelectromech. Syst. **5**, 60 (1992).
2. Mei, Y., Solovev, A.A., Sanchez, S. & Schmidt, O.G. Rolled-up nanotech on polymers: from basic perception to self-propelled catalytic microengines. Chem. Soc. Rev. **40**, 2109 (2011).
3. Bandari, V.K. et al. A flexible microsystem capable of controlled motion and actuation by wireless power transfer. Nat. Electron. **3**, 172 (2020).
4. Lee, Y. et al. Si chiplet-controlled 3D modular microrobots with smart communication in natural aqueous environments, Sci. Robot. **10**, eadu6007 (2025).
5. Lee, Y. et al. Nano-biosupercapacitors enable autarkic sensor operation in blood. Nat. Commun. **12**, 4967 (2021).

Band-Structure Engineering of TiO_2 with Pt Single Atoms for Photocatalytic Nitrate Reduction to Ammonia

Hayoon Jung[1, 2]

[1]Department of Materials Science and Engineering, WW4-LKO, Friedrich-Alexander-Universität Erlangen-Nürnberg, Martensstraße 7, 91058 Erlangen, Germany
[2]Regional Centre of Advanced Technologies and Materials, Palacký University, Šlechtitelů 27, 78371 Olomouc, Czech Republic
hayoon.jung@fau.de

Abstract. TiO_2 is a benchmark photocatalyst and a widely used support for metal single-atom co-catalysts. However, its conventional n-type character and band-edge positions restrict its potential in single-atom photocatalysis. This contribution summarizes two studies that address these limitations by jointly engineering the TiO_2 host and atomically dispersed Pt co-catalysts. First, quantum-confined p-type TiO_2 nanotubes were developed, enabling direct photocatalytic nitrate reduction to ammonia without the need of hole scavengers. The incorporation of Pt single atoms further enhanced the photocatalytic activity and shifted the selectivity toward ammonia. Second, this concept was extended to ultrathin Ti-deficient $Ti_{0.87}O_2$ nanosheets decorated with Pt single atoms, delivering efficient and stable nitrate-to-ammonia photocatalysis. Collectively, these studies demonstrate how engineering the electronic structures of TiO_2 can expand its reaction scope while improving the activity and selectivity of single-atom photocatalysts.

Keywords: TiO_2, single-atom catalyst, photocatalytic nitrate reduction

1. Introduction

Harnessing solar energy for chemical reactions, such as hydrogen evolution, CO_2 reduction, and nitrate (NO_3^-) reduction, with efficient photocatalytic platforms has drawn significant interest in recent years.[1–5]In classic photocatalysis, semiconductorsare widely adopted as promising catalysts since they can generate excited charge carriers (i.e., electron–hole pairs) upon irradiation with light of energy exceeding their band gap.[6]TiO_2 is among the most established semiconductor photocatalysts because of its chemical stability, low toxicity, and broad applicability.[7–10] However, most TiO_2 materials suffer from sluggish kinetics of charge transfer at the interface between the catalyst and the reactant.[11] In this regard, noble metal, such as Pt, co-catalysts are often required to promote the charge transfer and photocatalytic reactions.[12,13]Among numerous forms of co-catalysts,single-atom (SA) catalysts have gathered significant attention due to their maximum atom utilization and exceptional support interaction.[14,15]

While considerable efforts have been made to establish Pt SAs-decorated TiO_2 materials for prominent solar-to-chemical energy conversion, the influence of the electronic or defect structures of TiO_2 on SA photocatalysis is largely overlooked. For example, one potential approach to alter the energy level of a semiconductor's band structure and thereby extend the thermodynamic limits of photocatalytic reactions is to utilize quantum confinement effects, which result in a band gap broadening. However, for TiO_2, the quantum confinement effect and its photocatalytic applications are rarely investigated due to the large effective mass of charge carriers, which severely restricts the reaction spectrum in titania.[16,17] Another key challenge with TiO_2 is that it is inherently an oxygen-deficient and n-type material with relatively low hole mobility, which can impede the overall redox reaction and photocatalytic efficiency.[18,19]

In the present contribution, two related studies are highlighted, which resolve these coupled problems through quantum-confined and p-type TiO_2 supports. First, a novel photocatalytic platform for photocatalytic nitrate reduction to ammonia was developed by anchoring Pt SAs to p-type TiO_2 nanotubes with quantum confinement (p-TiO_2 NTs).[20] Nitrate is one of the major pollutants in wastewater and the selective reduction of nitrate into a desired product (e.g., ammonia) is an intriguing strategy for water remediation and upcycling.[21–23] However, considering the reduction potential of nitrate to ammonia ($NO_3^- + 9H^+ + 8e^- \rightarrow NH_3 + 3H_2O$, −0.58 V *vs.* NHE at pH 7), a direct electron transfer is not thermodynamically possible with a conventional conduction band level of TiO_2.[23–25] In this regard, quantum confinement effect can provide a higher thermodynamic energy of electrons of TiO_2 materials, allowing the direct photocatalytic nitrate reduction to ammonia. In addition, p-type conductivity can offer enhanced hole mobility and contribute to the

facilitation of overall redox reactions. Following this context, Ti-deficient TiO_2 ultrathin nanosheets ($Ti_{0.87}O_2$ NSs) with a wider band gap and stronger p-type characteristics were also developed for Pt SA deposition and for further utilization in photocatalytic nitrate reduction and ammonia production.[26]

2. Experiments and Discussion

The first study[20] achieved scavenger-free photocatalytic reduction of nitrate to ammonia by integrating Pt SAs with quantum-confined p-TiO_2 NTs prepared through hydrothermal methods (Fig. 1). The nanotubes crystallized in the monoclinic TiO_2(B) phase and possessed an outer diameter of ~10 nm with a wall thickness of ~2.7 nm, while uniform incorporation of Na dopants (≈4–5 at%) induced p-type conductivity. Incident photon-to-electron conversion efficiency (IPCE) and diffuse reflectance spectroscopy (DRS) measurementsindicated a band gap of 3.5 eV, and Mott–Schottky analysis yielded a conduction band position of −0.7 V *vs.* NHE at pH 7. This band-edge upshift enabled by quantum confinement rendered nitrate reduction feasible, where it is thermodynamically impossible on conventional TiO_2. In addition, more pronounced hole mobility by p-type conductivity can facilitate the counter reaction of sluggish water oxidation, further promoting the overall redox photocatalytic reactions.The p-TiO_2 NTs were then decorated with Pt SAs by a reactive deposition method, forming two-fold oxygen-coordinated Pt SAs with a formal charge (δ+) around δ≈ 2 on the p-TiO_2 NTs (Pt SAs/p-TiO_2 NTs). The Pt SA deposition (optimal Pt concentration: 0.27 at%) markedly enhanced both the activity and selectivity for ammonia formation. In addition, by tuning nanotube thickness and Na doping, this study further identified quantum confinement and p-type conductivity as key factors governing the photocatalytic performance and selectivity of the Pt SAs. Specifically, Pt SAs/p-TiO_2 NTs outperformed Pt SAs on thicker NTs (band gap 3.4 eV), n-type NTs, P25, and Pt nanoparticles/p-TiO_2 NTs, while also delivering a high NH_3 selectivity of 62.96%, far exceeding those of the other catalysts and the bare p-TiO_2 NTs. Finally, this study also elucidated the key role of each constituent in which p-TiO_2 NTs primarily facilitated the nitrate-to-nitrite step, whereas Pt SAs played a dominant role in reducing nitrite to ammonia.

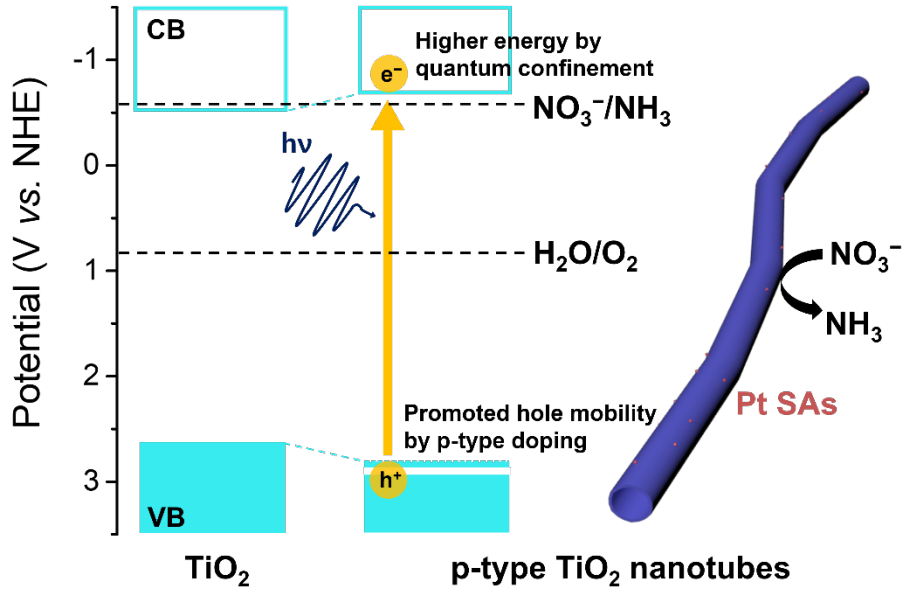


Fig. 1. A schematic illustration of photocatalytic nitrate reduction to ammonia in bulk TiO_2 and Pt SAs/p-TiO_2 NTs.

Similarly, the second study[26] reported the synthesis of Ti-deficient ultrathin $Ti_{0.87}O_2$NSs via protonation and delamination of lepidocrocite-type titanate, and their use for direct photocatalytic nitrate-to-ammonia conversion, with a substantial activity enhancement by introducing Pt SAs (Fig. 2). The resulting NSs exhibited an ultrathin morphology with lateral dimensions of several micrometers and a thickness of ~1.5 nm (bilayer), as confirmed by scanning electron microscopy (SEM), transmission electron microscopy (TEM), and rocking-curve analysis. X-ray diffraction (XRD) verified an interlayer spacing of 0.91 nm together with a characteristic anatase peak. Photoelectrochemical (IPCE), DRS, and Mott–Schottky analyses revealed a quantum-confinement-induced band gap of 3.7 eV and p-type conductivity (with higher hole density than that of p-TiO_2 NTs). The conduction band position was determined to be −0.8 V *vs.* NHE at pH 7 (more negative

than that of p-TiO_2 NTs), also enabling direct nitrate reduction to ammonia without any hole scavenger. Pt SAs were then anchored on the $Ti_{0.87}O_2$NSs by a simple dark deposition process; TEM confirmed atomic dispersion, while Pt 4f X-ray photoelectron spectra showed Pt^{4+} features for Pt SAs on $Ti_{0.87}O_2$NSs (Pt SAs/$Ti_{0.87}O_2$NSs). As a result, both activity and selectivity toward ammonia were significantly improved, outperforming Pt SAs/P25, Pt nanoparticles, the bare $Ti_{0.87}O_2$NSs, and P25 (Fig. 2). The optimal Pt loading was 0.49 at%, and the ammonia selectivity reached 64.54%, also far exceeding Pt SAs/P25, Pt nanoparticles/$Ti_{0.87}O_2$NSs, and the pristine NSs. The Pt SAs/$Ti_{0.87}O_2$NSs also maintainedtheir activity over 32 h with four consecutive cycles of photocatalysis.

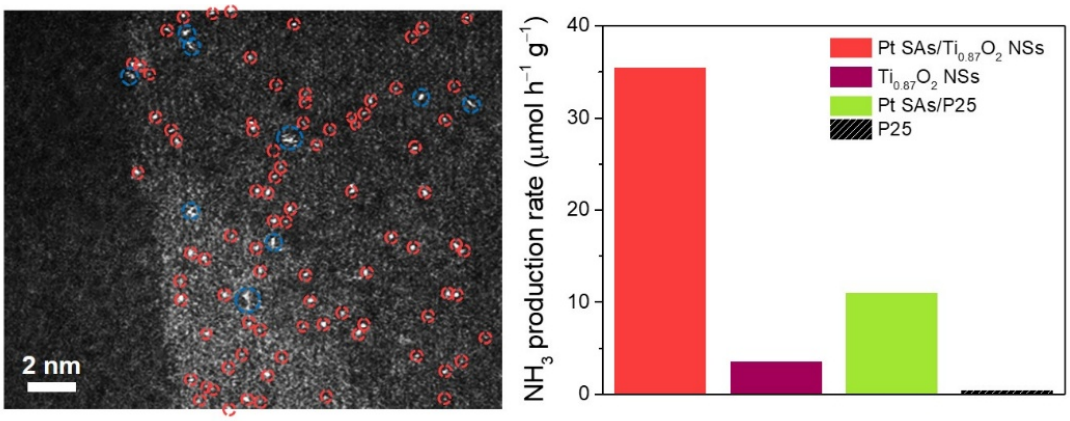


Fig. 2. TEM image of Pt SAs/$Ti_{0.87}O_2$ NSs (left) and photocatalytic activity of various catalysts for nitrate reduction to ammonia (right).

It is important to note that all photocatalytic reactions in these studies were carried out without the addition of any hole scavengers. In most TiO_2-based systems, with or without co-catalysts, hole scavengers such as methanol, humic acid, or formic acid are commonly employed for photocatalytic nitrate reduction because water oxidation is kinetically sluggish.[23] In addition, an aqueous radical species (e.g., $CO_2^{\bullet-}$),typically generated during the oxidation of these hole scavengers, is required in such reactions because direct electron transfer from the classic TiO_2 conduction band is thermodynamically unfavorable.[22–25]It is also worth mentioning that the aforementioned Pt SA-decorated TiO_2 photocatalysts in these studies may offer an advantage in selectively capturing the generated ammonia, as they suppress toxic nitrite formation and produce N_2 as the only side product, which is barely soluble in water.

3. Conclusions

In conclusion, this contribution reveals that engineering the band positions and conductivity types of TiO_2 provides an effective strategy for expanding its photocatalytic reaction scope and expediting SA photocatalysis. Quantum confinement in p-TiO_2 NTs and $Ti_{0.87}O_2$NSs widened the band gap and shifted the conductionband edge to a sufficiently negative potential for direct nitrate reduction to ammonia without the use of hole scavengers. The p-type conductivity of both nanostructures further improved the hole mobility, promoting the kinetically sluggish water oxidation. The incorporation of Pt SAs enhanced the photocatalytic activity and facilitated the multistep reduction pathway toward ammonia, resulting in higher activity and selectivity than that achieved with other reference catalysts. These findings highlight the synergistic roles of semiconductor band-structure engineering and atomically dispersed co-catalysts in designing efficient photocatalytic systems, particularly for water remediation and ammonia production.

References

1. Gao, C., Low, J., Long, R., Kong, T., Zhu, J., Xiong, Y.: Heterogeneous single-atom photocatalysts: fundamentals and applications. Chem. Rev. **120**(21), 12175–12216 (2020).
2. Roy, S.: Photocatalytic materials for reduction of nitroarenes and nitrates. J. Phys. Chem. C **124**(52), 28345–28358 (2020).
3. Wy, Y., Jung, H., Hong, J.W., Han, S.W.: Exploiting plasmonic hot spots in Au-based nanostructures for sensing and photocatalysis. Acc. Chem. Res. **55**(6), 831–843 (2022).

4. Lee, S.-U., Jung, H., Wi, D.H., Hong, J.W., Sung, J., Choi, S.-I., Han, S.W.: Metal–semiconductor yolk–shell heteronanostructures for plasmon-enhanced photocatalytic hydrogen evolution. J. Mater. Chem. A **6**(9), 4068–4078 (2018).
5. Jung, H., Whang, Y., Han, S.W.: Ag–CdS yolk–shell heteronanostructures for plasmon-enhanced photocatalysis. Bull. Korean Chem. Soc. **42**(5), 806–809 (2021).
6. Serpone, N., Emeline, A. V.: Semiconductor photocatalysis—past, present, and future outlook. J. Phys. Chem. Lett. **3**(5), 673–677 (2012).
7. Fujishima, A., Honda, K.: Electrochemical photolysis of water at a semiconductor electrode. Nature **238**(5358), 37–38 (1972).
8. Hamad, S., Catlow, C. R. A., Woodley, S. M., Lago, S., Mejías, J. A.: Structure and stability of small TiO_2 nanoparticles. J. Phys. Chem. B **109**(33), 15741–15748 (2005).
9. Jung, H., Song, J., Lee, S., Lee, Y. W., Wi, D. H., Goo, B. S., Han, S. W.: Hierarchical metal–semiconductor–graphene ternary heteronanostructures for plasmon-enhanced wide-range visible-light photocatalysis. J. Mater. Chem. A **7**(26), 15831–15840 (2019).
10. Jung, H., Cho, Y., Kang, S., Nho, H.-W., Kim, Y., Kwon, O.-H., Han, S. W.: Upconversion material–plasmonic metal–semiconductor ternary heteronanostructures for wide-range solar-to-chemical energy conversion. ACS Appl. Mater. Interfaces **16**(2), 2341–2350 (2024).
11. Guo, Q., Zhou, C., Ma, Z., Yang, X.: Fundamentals of TiO_2 photocatalysis: concepts, mechanisms, and challenges. Adv. Mater. **31**(50), 1901997 (2019).
12. Kawai, T., Sakata, T.: Photocatalytic hydrogen production from liquid methanol and water. J. Chem. Soc., Chem. Commun. **16**(15), 694–695 (1980).
13. Yang, J., Wang, D., Han, H., Li, C.: Roles of cocatalysts in photocatalysis and photoelectrocatalysis. Acc. Chem. Res. **46**(8), 1900–1909 (2013).
14. Wu, S.-M., Schmuki, P.: Single atom cocatalysts in photocatalysis. Adv. Mater. **37**(7), 2414889 (2025).
15. Jung, H., Cha, G., Kim, H., Will, J., Zhou, X., Bad'ura, Z., Zoppellaro, G., Dobrota, A. S., Skorodumova, N. V., Pašti, I. A., Sarma, B. B., Schmidt, J., Spiecker, E., Breu, J., Schmuki, P.: Cation vacancies in Ti-deficient TiO_2 nanosheets enable highly stable trapping of Pt single atoms for persistent photocatalytic hydrogen evolution. Small **21**(29), 2502428 (2025).
16. Bavykin, D. V., Walsh, F. C.: Titanate and titania nanotubes: synthesis, properties and applications. Royal Society of Chemistry, Cambridge (2009).
17. Monticone, S., Tufeu, R., Kanaev, A. V., Scolan, E., Sanchez, C.: Quantum size effect in TiO_2 nanoparticles: does it exist? Appl. Surf. Sci. 162–163, 565–570 (2000).
18. Yang, S., Lu, Q., Wang, F., Zhi, Y., Chen, J., Wang, Y., Zhang, H., Yin, H., Sun, P., Cao, W.: S-scheme SnO/TiO_2 heterojunction with high hole mobility for boosting photocatalytic degradation of gaseous benzene. Chem. Eng. J. **478**, 147345 (2023).
19. He, Z.-K., Lu, Y., Zhao, J., Zhao, J., Gao, Z., Song, Y.-Y.: Engineering carrier density at TiO_2 nanotube metasurface with hole reservoir for enhanced photo-electrocatalysis. Appl. Surf. Sci. **613**, 155974 (2023).
20. Jung, H., Kim, H., Will, J., Spiecker, E., Schmuki, P.: p-Type TiO_2 nanotubes: quantum confinement and Pt single atom decoration enable high selectivity photocatalytic nitrate reduction to ammonia. Angew. Chem. Int. Ed. **64**(22), e202415865 (2025).
21. Youn, G., Hong, I., Song, B., Lee, D., Yun, Y., Moon, H. S., Yong, K.: Present state and future outlook of ammonia production through photocatalytic nitrate reduction. Sol. RRL **8**(3), 2300880 (2024).
22. Hirakawa, H., Hashimoto, M., Shiraishi, Y., Hirai, T.: Selective nitrate-to-ammonia transformation on surface defects of titanium dioxide photocatalysts. ACS Catal. **7**(5), 3713–3720 (2017).
23. Tugaoen, H. O. N., Garcia-Segura, S., Hristovski, K., Westerhoff, P.: Challenges in photocatalytic reduction of nitrate as a water treatment technology. Sci. Total Environ. 599–600, 1524–1551 (2017).
24. Bems, B., Jentoft, F. C., Schlögl, R.: Photoinduced decomposition of nitrate in drinking water in the presence of titania and humic acids. Appl. Catal. B Environ. **20**(2), 155–163 (1999).
25. Xiong, Y., Wang, Y., Zhou, J., Liu, F., Hao, F., Fan, Z.: Electrochemical nitrate reduction: ammonia synthesis and the beyond. Adv. Mater. **36**(17), 2304021 (2024).
26. Jung, H., Cha, G., Kim, H., Will, J., Zhou, X., Spiecker, E., Breu, J., Schmuki, P.: Ultrathin Ti-deficient TiO_2 nanosheets with Pt single atoms enable efficient photocatalytic nitrate reduction to ammonia. J. Am. Chem. Soc. **147**(11), 9049–9055 (2025).

Accelerating Alloy Design using AI: It's Optimization All the Way Down

Jin-Young Lee[1]

[1]Max Planck Institute for Sustainable Materials, Max-Planck-Straße 1, 40237 Düsseldorf, Germany
jinyoung.lee@mpi-susmat.de

Abstract. Alloy design, at its core, is an optimization problem: finding the composition and processing conditions that satisfy a set of target properties. This paper argues that AI-driven alloy design does not remove this problem, but distributes it across multiple nested levels, each itself an optimization task. Training a surrogate model optimizes its internal parameters against prediction error, while tuning its hyperparameters is a second-order optimization to avoid overfitting issue. Searching an intractably large compositional space is an optimization of where to sample next, formalized through acquisition functions in Bayesian optimization or through population-based search such as genetic algorithms. Balancing competing target properties becomes a multi-objective optimization, addressed through Pareto front analysis, and selecting a small number of candidates for costly experimental validation is itself an optimization of coverage under a limited budget, approximated here through K-means clustering. These ideas are illustrated through a case study on medium-Mn steels, whose predictions guided a multi-objective search over roughly 17.9 million candidates, ultimately yielding five experimentally validated alloys exceeding 2 GPa tensile strength.

Keywords: Alloy Design, Machine Learning, Optimization, Inverse Design

1. Introduction

With the rapid growth of open-source machine learning (ML) frameworks and user-friendly software, researchers without a formal computer science background have increasingly been able to apply AI techniques to their own domains. In materials science, this has led to two broad lines of work: image-based microstructure analysis, and regression-based prediction of material properties from composition and processing variables.

This paper focuses on the second line of work, and specifically on a problem that is often underestimated: building an accurate predictive model is only the first step, and does not by itself solve the alloy design problem. Even after a model has been trained to predict target properties with high accuracy, finding the composition and processing conditions that achieve a desired property combination remains difficult, because the space of possible compositions and processing routes is far too large to search exhaustively.

This paper argues that this difficulty is structural, not incidental: alloy design, whether performed with or without AI, is fundamentally an optimization problem, namely finding the input conditions that satisfy a set of target output properties. What AI changes is not the presence of this optimization problem, but its structure. Instead of one large, intractable optimization problem, AI-driven alloy design decomposes the task into a series of smaller, more tractable optimization problems, each occurring at a different stage of the workflow, from data preparation, to model training, to candidate search, to final candidate selection for experimental validation.

To illustrate this structure concretely, this paper uses case study on medium-Mn steels [1-3], aiming to identify alloy compositions and processing conditions that simultaneously improve tensile strength and resistance to hydrogen embrittlement.

2. Discussion

Preparing the Ground: Data Collection and Curation

Before any optimization can take place, a dataset connecting alloy composition and processing conditions to target properties must exist. This stage is not itself an optimization problem, but it determines the quality of every optimization problem that follows.

In this case study, approximately 1,600 data points covering 36 input features (chemical composition and processing conditions) and two output properties (ultimate tensile strength (UTS) and total elongation (TE)),

were manually collected from ~900 peer-reviewed papers on medium-Mn steels published between 2012 and 2024 [1,2]. This manually curated dataset was then used as ground truth to develop and validate an automated extraction pipeline based on a large language model (LLM), converting unstructured text into structured data.

Two practical difficulties arose. First, values reported only in figures, rather than in text or tables, remain difficult to extract automatically. Second, processing conditions are frequently categorical rather than numerical, for example, whether a sample was air-cooled, water-quenched, or furnace-cooled. Such variables must be mapped to a consistent numerical encoding (e.g., 1 for air cooling, 2 for quenching, 3 for furnace cooling) before use. Because this encoding has no physical meaning by itself, the specific choice of encoding scheme must be explicitly defined and communicated to the extraction pipeline through careful prompt engineering.

The manual literature search described above, while necessary to build a validated ground-truth dataset, is itself becoming increasingly automatable. Publisher and database APIs (e.g., Elsevier, Springer) can now be used to programmatically retrieve papers matching a given query and feed them directly into the LLM-based extraction pipeline, assembling literature retrieval and data extraction into a single automated workflow and considerably reducing the manual effort required as new papers are published.

Optimizing Machine Learning Model: Parameters and Hyperparameters

Once the dataset is available, the first optimization problem is building an accurate surrogate model, at two distinct levels: the model's internal parameters, and its hyperparameters.

Parameter-level optimization. Training a machine learning model means finding the internal parameters that minimize prediction erroron the training data. The choice of model architecture strongly affects how well this succeeds. Neural networks are often the default choice when many correlated inputs are involved, and perform well when inputs are numerical and physically continuous. In this case, however, the dataset contained a mixture of continuous features (composition, temperature) and categorical features (cooling method), encoded as integers with no inherent physical ordering. Because neural networks treat all inputs as continuous quantities, they were misled by this arbitrary encoding. Tree-based ensemble models, which make decisions through recursive binary splits (e.g., "Mn content above 7.5 wt%?" or "air-cooled or not?"), were far less sensitive to this encoding and consistently outperformed neural networkson this dataset [1,2,4].

Hyperparameter-level optimization. Beyond the model's internal parameters, its configuration (e.g., tree depth, number of estimators, learning rate) must also be optimized. This optimization, however, must be approached with caution: if the model is tuned to fit the training data too closely, it begins to memorize noise and sample-specific details rather than learning the underlying trend, and its performance on unseen test data deteriorates as a result. Hyperparameter optimization must therefore explicitly guard against overfitting the training data, rather than simply minimizing training error. In this case, K-fold cross-validation was used for this purpose: the training data was split into K folds, with K−1 folds used for training and the remaining fold used for validation, repeated K times with different fold assignments. The resulting average validation performance was used to guide hyperparameter selection, while the test set was kept completely separate and untouched throughout this process, ensuring that the reported test performance reflects genuine generalization rather than an artifact of hyperparameter selection. This procedure allowed both the training and test accuracy to be raised to a satisfactory level, and the balance between them can be summarized as an overfitting index, defined as the ratio of test accuracy to training accuracy: a value close to 1 indicates that the model generalizes well, whereas a substantially lower value indicates that the model has overfit the training data.

Defining the Search Space

Even with an accurate surrogate model, alloy design is not solved: the space of possible compositions and processing conditions is vast. In this case study, the goal was to identify compositions and processing conditions that simultaneously improve tensile strength and hydrogen embrittlement (HE) resistance. Rather than allowing all 36 input features to vary freely, the search space was first constrained using domain knowledge and ab-initio calculations. Micro-alloying elements were screened for their effect on hydrogen diffusivity and grain boundary cohesion, since intergranular fracture along grain boundaries is the dominant HE failure mode in high-strength steels [3]. Composition ranges for key elements (e.g., Mn and C) were bounded based on known their trade-offs with cold-rollability [1]. Processing variables with no established effect on the target properties were fixed rather than optimized. This step is not itself an optimization algorithm, but it determines the feasible region over which the subsequent optimization is performed. Even after this reduction, the remaining search space still contained approximately 17.9 million candidate combinations, far too large to evaluate exhaustively. A grid search over this space would require several days of continuous computation, even if each candidate could be evaluated in milliseconds.

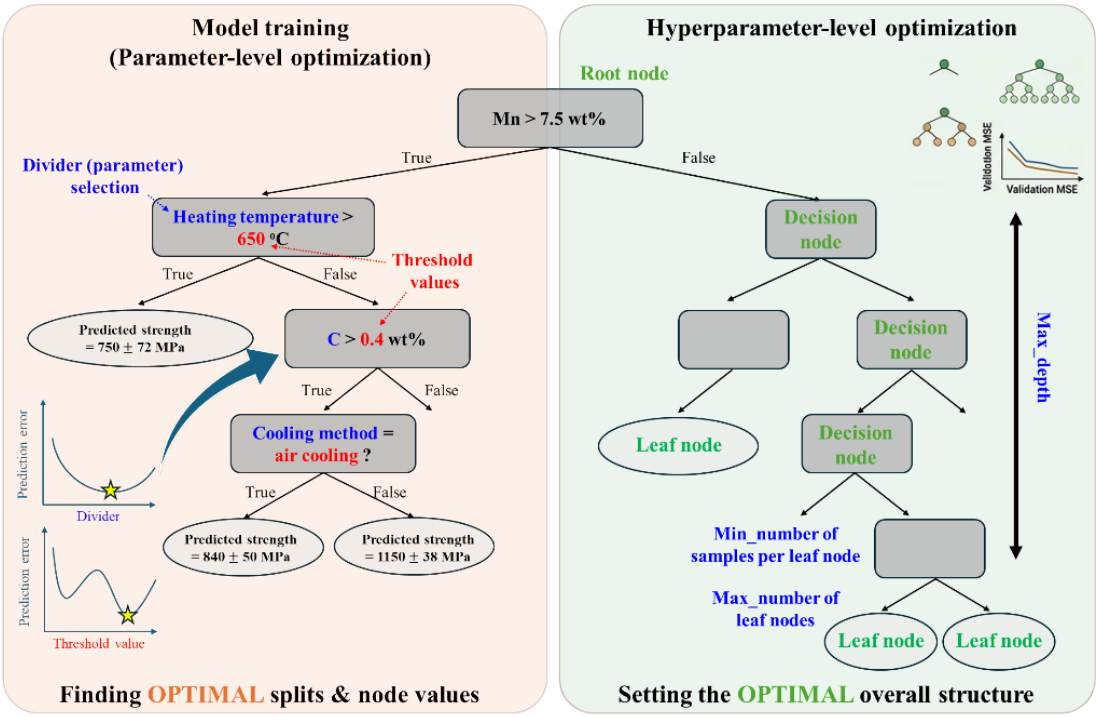


Fig. 1. Dual-stage optimization in a regression tree model. As an example, model training (left) optimizes internal parameters (split thresholds and leaf values) to minimize training MSE, while hyperparameter optimization (right) selects the optimal model structure (e.g., tree depth and count) using validation performance.

Searching the Space: Bayesian Optimization and Genetic Algorithms

Two conceptually different strategies were used to search this reduced space, each optimizing a different quantity.

Bayesian optimization (BO). Any search strategy must balance exploration against exploitation. Exploration means sampling regions where the model is uncertain, in order to improve it. Exploitation means sampling regions the model already predicts to be promising, in order to refine the best solution found so far. BO formalizes this trade-off explicitly [5]. A Gaussian Process surrogate is fitted to the available data. This provides not only a predicted value at any point in the search space, but also an estimate of the uncertainty of that prediction (Fig. 2a). An acquisition function, such as the Upper Confidence Bound (UCB) or Expected Improvement (EI), combines these two quantities into a single score. The next candidate to evaluate is then chosen by optimizing this acquisition function. BO is therefore optimization used to decide where the next optimization step should be performed.

This structure extends naturally to multiple objectives. Because tensile strength and HE resistance could not be improved independently, candidates were evaluated in terms of the Pareto front, the set of non-dominated solutions for which improving one property requires sacrificing the other. At each iteration, one candidate was selected from the Pareto front with the highest predictive uncertainty, favoring exploration and improving the model where it is least reliable. A second candidate was selected with the highest dominance count, the number of other candidates it outperforms in both objectives simultaneously, favoring exploitation and reinforcing the strongest known trade-offs (Fig. 2b). Iterating in this way simultaneously expanded the Pareto front and reduced model uncertainty across it.

Genetic algorithms (GA). When the goal is instead to identify as many diverse, high-performing candidates as possible according to the surrogate's predictions, population-based metaheuristics such asGA or particle swarm optimization (PSO) provide an efficient alternative. In a GA, each candidate composition and processing condition is encoded as a string. A population of such candidates is then evolved over successive generations through three operations [7]. Selection retains the best-performing candidates. Crossover combines features of two parent candidates. Mutation randomly alters some features to maintain diversity. Unlike BO, this process does not explicitly reason about uncertainty, but it explores the surrogate's prediction landscape efficiently and in parallel. Applied to this search space, the GA identified a large number of candidate compositions with predicted UTS exceeding 2 GPa.

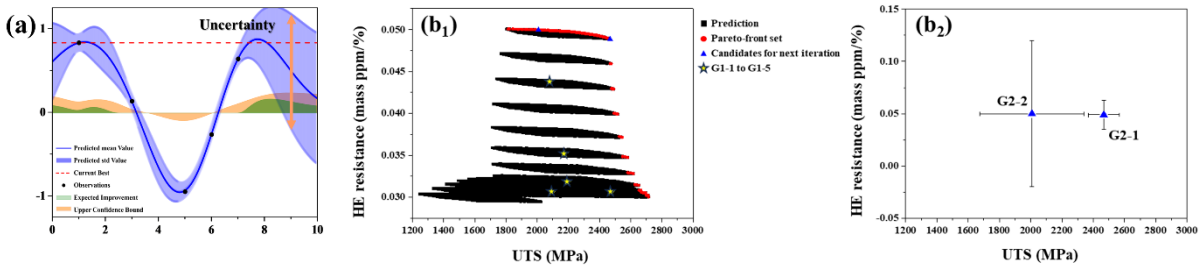


Fig. 2. Conceptual framework and multi-objective execution of Bayesian Optimization (BO) for alloy design. (a) Schematic illustration of Gaussian Process regression and acquisition functions (UCB and EI), balancing exploration (high predictive uncertainty) and exploitation (high predicted performance) [6]. (b_1) Multi-objective optimization searching space balancing UTS and HE resistance, where candidates for the next iteration (blue triangles) are selected from the Pareto front (red dots) based on predictive uncertainty and dominance count. (b_2) Property evaluation and predictive uncertainty of selected candidate alloys (e.g., G2-1 and G2-2) [3].

Selecting Candidates for Experimental Validation using Unsupervised Learning

After the search algorithms converged, a large number of promising candidates remained, still far more than could realistically be fabricated and tested. The final step is to choose a small number of candidates that represent this larger pool as well as possible. The goal here is to optimize the coverage of the promising region of the search space under a fixed experimental budget.

In this case study, all candidates predicted to exceed 2 GPa UTS by the GA were grouped using K-means clustering with K equal to five. For each cluster, the average chemical composition and austenitizing temperature were taken as a representative candidate. This approach favors diversity over the search space, rather than favoring only the single best-predicted point, and reduces the risk that all fabricated alloys fail for the same unforeseen reason. All five alloys fabricated and tested according to these cluster centroids exceeded 2 GPa UTS experimentally, confirming both the surrogate model and the search strategy used to generate the candidate pool.

3. Conclusions

This paper has walked through an AI-driven alloy design workflow, from data curation to experimental validation, using medium-Mn steels as a case study. At every stage, a different quantity was optimized: model parameters against prediction error, hyperparameters against validation error, the next candidate to evaluate via acquisition functions or population dynamics, a Pareto front rather than a single value, and finally experimental coverage under a limited budget. None of these problems disappeared with AI; a single intractable optimization problem was instead decomposed into smaller, tractable ones, each solved with a tool suited to its structure.

References

1. J.-Y. Lee, Y.-K. Lee *et al*., Design of high strength medium-Mn steel using machine learning, *Materials Science and Engineering A*, 843 (2022) 143148
2. J.-Y. Lee, Y.-K. Lee *et al*., Inverse design of high strength medium-Mn steel using a machine-learning-aided genetic algorithm, *Journal of Materials Research and Technology*, 33 (2024) 2672-2682.
3. J.-Y. Lee, Y.-K. Lee *et al*., Design of hydrogen-resistant high-strength medium-Mn steels *via* Pareto active learning, *Acta Materialia 312 (2026) 122264.*
4. L. Grinsztajn, E. Oyallon, G. Varoquaux, Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems* (2022) 35.
5. J. Mockus, Bayesian approach to global optimization, Mathematics and itsapplications. Kluwer Academic; 1989.
6. S.-H. Kim, J.-Y. Lee, Y.-K. Lee, A comparative study of query strategies in Pareto active learning for heat treatment optimization of medium-Mn steel, *Journal of Materials Research and Technology*, 37 (2025) 2152-2164.
7. A.E. Eiben, C.A. Schippers, On evolutionary exploration and exploitation, *FundamentaInformaticae*, 35 (1998) 35-50.

Electrical, Electronics & Information Technology Division

Welcoming Message

As a specialized division of VeKNIe.V., the Electrical, Electronics & Information Technology Division promotes scientific and technological advancement across electrical engineering, electronic systems, information technology, bio-integrated engineering, display technology, and photonics. The Division welcomes high-quality research papers, technical reports, review articles, and industrial case studies that address both established disciplines and emerging technologies. It provides a rigorous platform for sharing impactful research, fostering interdisciplinary collaboration, and translating innovative ideas into intelligent, connected, and sustainable solutions.

Focus & Core Scope

- **Electronics & Semiconductor Technology**
  Analog/mixed-signal ICs, RF/mmWave circuits, semiconductor devices, sensors, and advanced packaging.

- **Information Technology & Computing**
  Computer systems, distributed computing, cybersecurity, cloud platforms, and data engineering.

- **Artificial Intelligence & Intelligent Systems**
  Machine learning, edge AI, intelligent control, and data driven engineering.

- **Communication & Networking**
  Wireless systems, 5G/6G, IoT networks, and signal processing.

- **Bio-information & Bio-Integrated Engineering**
  Biomedical sensors, biosignal processing, bioelectronics, neural interfaces, and computational biology.

A 22 nm FD-SOI CMOS Fully Differential Two Stage Folded-Mesh CLASS AB Operational Amplifier with an Internal Regulator

Jeongwook Koh[1]

[1]Renesas Electronics Germany GmbH., Grenzstraße 28, 01109 Dresden, Germany
jeongwook.koh@vekni.org

Abstract. In this paper, we propose a 22nm FD-SOI (Fully Depleted Silicon-On-Insulator) CMOS fully differential two-stage folded-mesh CLASS AB operational amplifier with an internal regulator. The proposed amplifier is employed in a switched capacitor amplifier implementation. The performances of the switched capacitor amplifier are presented. The operation amplifier has open-loop DC gain of 147 dB, unity gain bandwidth of 59 MHz, and phase margin of 68 degree at 5-pF load in simulation. 50 dB power supply rejection on the outputs is achieved at 100 kHz, thanks to the internal regulator. The switched capacitor amplifier has voltage gain of8 V/V, and FFT (Fast-Fourier Transform) spectrum simulation yields SNDR (Signal-to-Noise and Dynamic Range) of 78.2 dB and SFDR (Spurious-Free Dynamic Range) of 78.6 dB, when a 95 mVpeak sinusoidal signal with 12.945 kHz is applied. The switched capacitor amplifier consumes 2.45 mW at sampling clock of 104.2 kHz and occupies an active die area of 0.0946 mm^2.

Keywords: 22nm FDSOI CMOS, folded-mesh, class AB, operational amplifier, switched capacitor amplifier.

1. Introduction

Switched capacitor (SC) circuits constitute fundamental building blocks in modern mixed-signal System-on-Chip (SoC) architectures, finding extensive applications in analog-to-digital converters (ADCs), digital-to-analog converters (DACs), sampled-data filters, and sensor interface circuits. In sample-and-hold systems and SC amplifiers, the operational amplifier (op-amp) acts as the core active component determining overall speed, linearity, dynamic range, and power dissipation. A key requirement in high-precision SC implementations is rapid and accurate transient settling, where the output signal must settle within a tight error margin (e.g., 0.01% or 12 to 16 bits of resolution) within a fraction of the clock phase [1].

Conventional Class-A operational amplifiers are intrinsically limited in slew rate and output driving capacity by their constant tail currents. Achieving fast settling in Class-A amplifiers under large capacitive loads necessitates high quiescent currents, severely compromising energy efficiency. Conversely, Class-AB operational amplifiers offer a superior power-to-bandwidth tradeoff by providing adaptive push-pull current drive. Under transient dynamic conditions, a Class-AB output stage can supply output currents significantly exceeding its quiescent current, thereby eliminating slew-rate bottlenecks while maintaining low static power consumption during steady-state conditions [2-5].

In nanoscale CMOS nodes, particularly below 45 nm, low supply voltages (typically 0.8 V to 1.0 V) impose severe voltage headroom limitations on traditional Class-AB control techniques. Legacy architectures, such as floating voltage source Class-AB control, suffer from reduced transistor voltage headroom, increased signal distortion, and heightened sensitivity to process, voltage, and temperature (PVT) variations. To overcome these challenges, folded-mesh Class-AB topologies incorporating minimum current selectors have emerged as an effective methodology for ultra-low-voltage mixed-signal applications. Furthermore, integrating high-density analog cores into mixed-signal SoCs exposes sensitive analog nodes to substantial digital power supply noise and ripple. An integrated power regulator is essential to reject high-frequency supply interference and maintain high dynamic range without requiring external dedicated supply rails.

This paper presents the detailed analysis, architecture, design equations, and physical implementation of a 22 nm FD-SOI CMOS fully differential two-stage folded-mesh Class-AB operational amplifier with an

integrated low-dropout (LDO) regulator. The paper is organized as follows: Section 2 details the operating principles of the folded-mesh Class-AB architecture with minimum current selection. Section 3 describes the full operational amplifier topology, includinginput stage, regulator integration, and common-mode feedback loops. Section 4 presents the switched capacitor amplifier application, layout implementation, and dynamic performance. Section 5 summarizes simulation results and provides comparative evaluations. Section 6 concludes the paper.

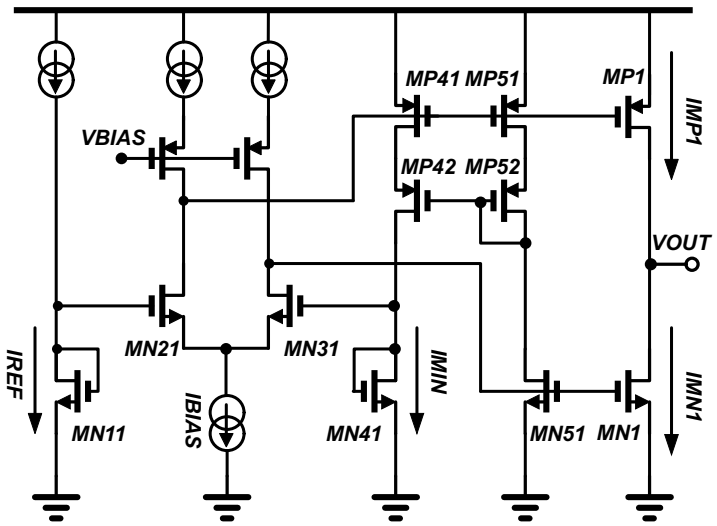


Fig. 1. Folded-mesh class AB with minimum current selector.

2. Principle: Folded-Mesh Class AB with Minimum Current Selector

The control loop of a Class-AB output stage must fulfill two fundamental requirements: (1) prevent the output transistors from turning completely off during large transient excursions to eliminate turn-on delay and crossover distortion, and (2) strictly establish and stabilize the quiescent bias current (IQ) in steady state regardless of PVT fluctuations.

Figure 1 illustrates the schematic concept of the proposed folded-mesh Class-AB output stage integrated with a minimum current selector. The output branch comprises PMOS driver MP1 and NMOS driver MN1. The control network consists of folded-mesh transistors (MN11, MN21, MN31, MN41), minimum current selector devices (MP41, MP42, MP51, MP52, MN51), and negative feedback bias loops. The quiescent current IQ of the outputs is set to the reference current IREF by folded mesh (MN11, MN21, MN31, MN41), minimum current selector (MP41, MP42, MP51, MP52, MN51) and output transistors (MP1, MN1) and their negative feedback loops.

When the output transistor MP1 starts to “push” some currents to a load, the current IMN1 of MN1 is selected as the minimum current IMIN. Now the minimum current IMIN is set to the reference current IREF. When the output transistor MN1 starts to “pull” some currents from the load, the current IMP1 of MP1 is chosen as the minimum current IMIN, which the minimum current IMIN is set to the reference current IREF.

Under quiescent conditions, the closed negative feedback loop forces the quiescent current IQ flowing through MP1 and MN1 to track the reference current IREF. When a large differential input voltage drives the amplifier stage:

- Push Mode Operation: Transistor MP1 is driven strongly into conduction to push substantial current into the load. Concurrently, the current IMN1 of NMOS driver MN1 decreases. The minimum current selector detects this reduction and clamps IMN1 to a predetermined minimum current IMIN, which is equalized to IREF through negative feedback.

- Pull Mode Operation: Transistor MN1 is driven strongly to sink heavy current from the load. The current IMP1 of PMOS driver MP1 drops until it is selected as the minimum current IMIN. The minimum current loop holds IMP1 equal to IREF, preventing MP1 from turning off.

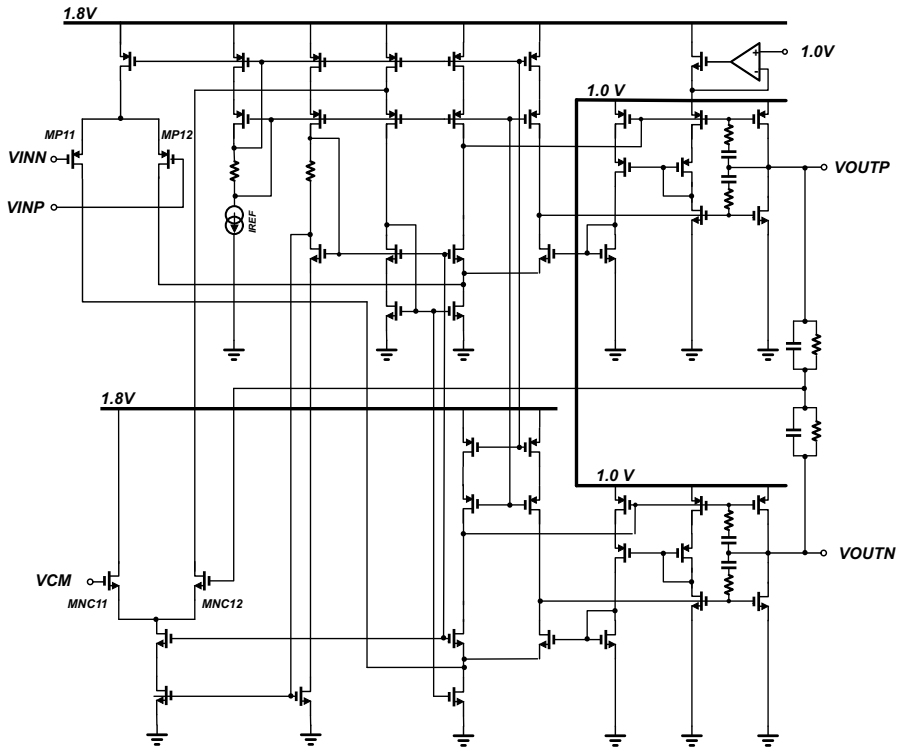


Fig. 2. Folded-mesh class AB with minimum current selector.

A critical advantage of the proposed folded-mesh Class-AB topology is its low minimum power supply requirement. The theoretical minimum supply voltage *VDD, min* for the output stage is expressed as:

$$V_DD,min = V_SG,MP1 + V_DSAT,MN21 + V_DSAT,IBIAS \quad (2)$$

Because the folded mesh avoids stacking multiple gate-source voltage drops between the supply rails, the circuit operates reliably down to sub-1.0 V supply levels, making it ideally suited for advanced 22 nm FD-SOI CMOS processes where supply voltage scaling is essential.

3. Application of the principle to Fully Differential Folded-Mesh Class AB Operational Amplifier

In this section, we present a fully differential folded-mesh Class AB amplifier with an internal regulator in Figure 2. It consists of a PMOS folded cascade (MP11, MP12), two folded-mesh class AB with minimum current selectors, an internal regulator, a current-mode common-mode feedback (MNC11, MNC12) and bias circuits. The detailed schematic architecture of the fully differential two-stage operational amplifier with an internal regulator is structured to maximize DC gain, bandwidth, power supply rejection, and common-mode stability.

In the design, the internal regulator regulates 1.8V power supply to generate 1.0V internal power supply. The minimum current selectors and output transistors are designed to operateat the 1.0V internal power supply. Current-mode common-mode feedback is employed for its faster settling of the common-mode feedback loop.

3.1 First-Stage PMOS Folded Cascode Architecture

The first stage employs a PMOS differential folded-cascode configuration (MP11, MP12). PMOS input devices are selected to minimize flicker noise (1/f noise) and accommodate appropriate input common-mode voltage ranges. The folded-cascode topology provides high output impedance, contributing significantly to the overall open-loop DC gain of 147 dB while allowing sufficient output signal swing to drive the second stage folded-mesh Class-AB control gates.

3.2 Second-Stage Folded-Mesh Class-AB Output & CMFB Subsystem

To implement fully differential signaling, two distinct folded-mesh Class-AB control networks with minimum current selectors are utilized—one for each differential output branch. Fully differential operational amplifiers require active Common-Mode Feedback (CMFB) to stabilize output common-mode voltages. In this design, a current-mode CMFB circuit (MNC11, MNC12) is integrated into the core topology. Current-mode CMFB is selected over conventional voltage-mode sampling due to its superior bandwidth, minimal phase lag, and faster dynamic settling of common-mode disturbances.

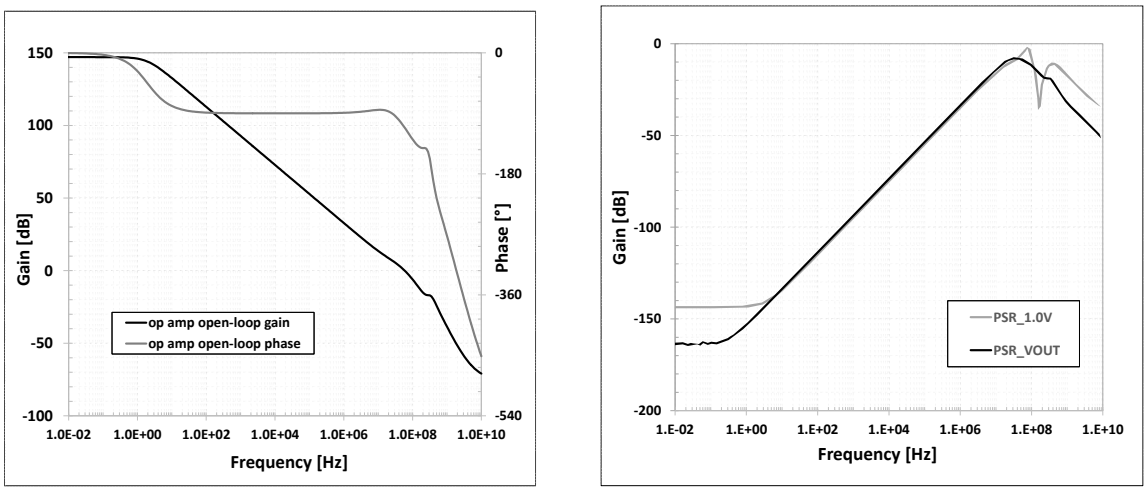


Fig. 3. Open-loop ac characteristics of the operational amplifier (left) and 1.8 V power supply rejection on the internal 1.0 power supply and the output(right)

3.3 Integrated Internal Power Supply Regulator

Nanoscale baseline microelectronic processes often operate with an external 1.8 V I/O supply rail, whereas thin-oxide core devices require 1.0 V or below for reliability and breakdown prevention. An on-chip low-dropout (LDO) regulator is embedded directly within the operational amplifier macro. The regulator converts the incoming 1.8 V power supply down to a ripple-free 1.0 V internal supply rail that powers the minimum current selectors and output stages.

This integrated regulator serves two primary purposes: (1) it ensures voltage compatibility and long-term reliability for thin-oxide 22 nm FD-SOI devices, and (2) it acts as a low-pass supply noise barrier, improving power supply rejection ratio (PSRR). At 100 kHz, the combined PSRR of the 1.0 V internal node and the differential outputs achieves 50 dB. The total static current consumption of the operational amplifier core under a 1.8 V supply is 388 µA.

3.4 Design Results

Figure 3 shows the open-loop ac characteristics of the proposed operational amplifier in the left, where1.8 V power supply rejection characteristics on 1.0 V internal power supply and the outputs of the proposed operational amplifier are also shown in the right.

Open-loop dc gain of 147 dB, unity-gain bandwidth of 59 MHz and phase margin of 68 degree are obtained at 5 pF load by simulation. Figure 3-(b) shows 1.8 V power supply rejection characteristics on 1.0 V internal power supply and the outputs of the proposed operational amplifier. In the simulation, power supply rejection of the 1.0 V internal power supply and the outputs achieves 50 dB at 100 kHz. The proposed amplifier consumes 388 uA at 1.8V.

4. A Switched capacitor amplifier with the proposed class AB operational amplifier

A fully differential switched capacitor (SC) gain stage was designed (Figure 4 (a)) and evaluated fordemonstration of the practical efficacy of the proposed Class-AB operational amplifier in a realistic mixed-signal environment.

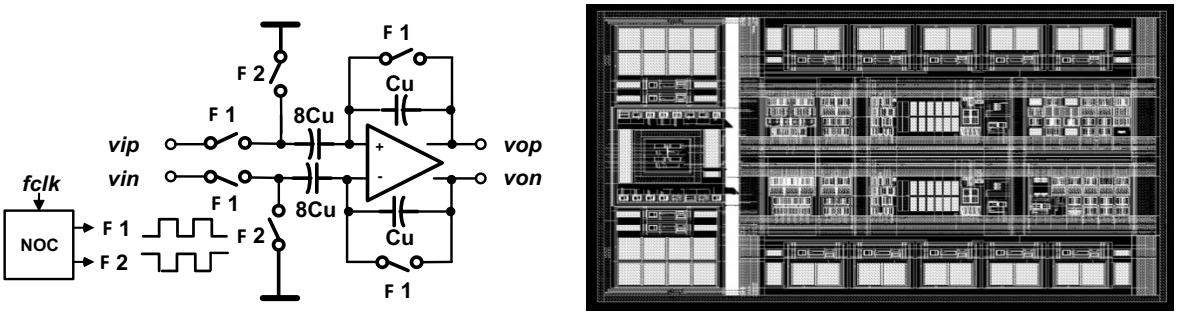


Fig. 4. The switched capacitor amplifier with the proposed class AB amplifier (a) and its layout implmentation (b)

4.1 SC Amplifier Topology and Clocking Strategy

The SC amplifier configuration employs a charge-redistribution architecture designed for a closed-loop non-inverting voltage gain of 8 V/V. Non-overlapping two-phase clock signals (Φ_1 and Φ_2) are generated on-chip from an external master sampling clock of 104.2 kHz.

- **Sampling Phase (Φ1):** Input voltages are sampled onto input sampling capacitors CS while the operational amplifier feedback loops are reset to establish initial charge states.
- **Transfer Phase (Φ_2):** Sampled charge is transferred across feedback capacitors C_F, generating the output voltage $V_{OUT} = (C_S / C_F)*V_{IN}$

To minimize total silicon area and thermal kT/C noise while maintaining precise matching, a unit capacitor Cuof 0.5 pF is selected. Capacitor arrays are constructed using common-centroid layout patterns to mitigate process gradient errors.

4.2 Settling Behavior and Dynamic Range Performance

Since the proposed Class-AB op-amp delivers dynamic boosting current during large signal transitions, the SC amplifier exhibits rapid settling during Phase Φ_2 without requiring high continuous quiescent current [3]. Dynamic spectral purity was evaluated via Fast Fourier Transform (FFT) analysis on the sampled output

waveform under full-scale driving conditions.

When driven by a 95 mV_{peak} sinusoidal input at a frequency of 12.945 kHz under a sampling clock of 104.2 kHz, FFT spectral simulations yielded:

- **Signal-to-Noise and Distortion Ratio (SNDR):** 78.2 dB
- **Spurious-Free Dynamic Range (SFDR):** 78.6 dBc
- **Effective Number of Bits (ENOB):** ~12.7 Bits

The complete switched capacitor amplifier consumes 2.45 mW from the 1.8 V power supply at the maximum clocking rate of 104.2 kHz (Figure 5).

5. Layout Implementation & Silicon Area Analysis

Physical implementation was carried out in a 22 nm FD-SOI CMOS fabrication process (Figure 4). Ultra-thin body and buried oxide (UTBB) features of 22 nm FD-SOI technology provide effective back-gate biasing capabilities and reduced parasitic drain/source capacitance, which enhances amplifier speed and power efficiency.

Careful floor planning was executed to isolate high-speed digital switching clock lines from sensitive analog differential input traces and bias nodes. Dummy device insertion and symmetry guard rings were incorporated to preserve inter-device matching across process corners,

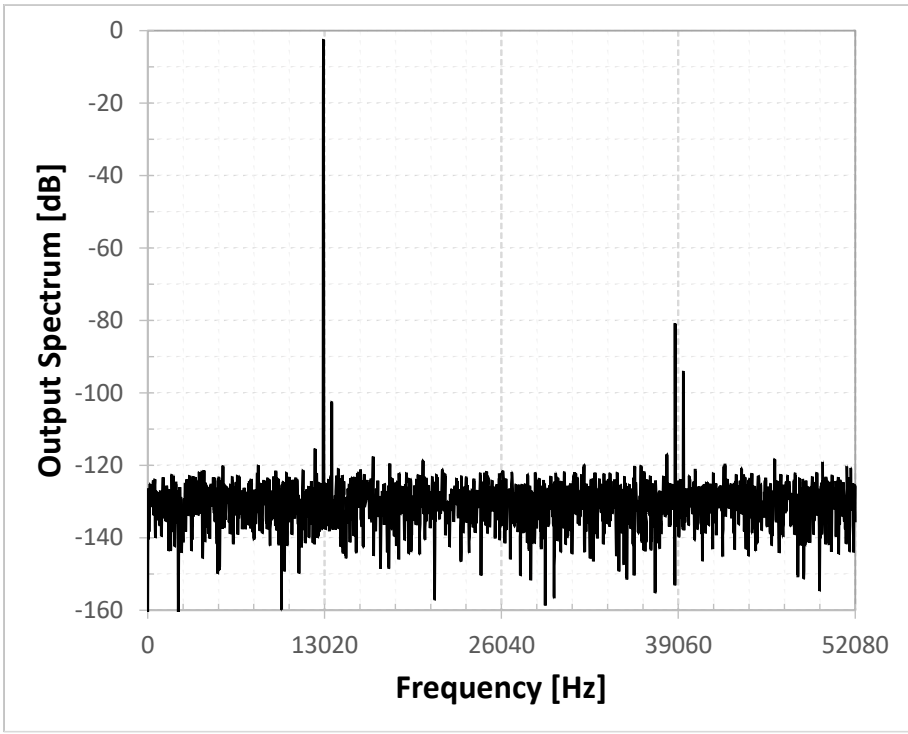


Fig. 5. FFT spectrum of the output signal of the switched capacitor amplifier.

6. Conclusions

In this paper, a 22 nm FD-SOI CMOS fully differential two-stage folded-mesh Class-AB operational amplifier with an integrated internal regulator has been presented and evaluated. By incorporating folded-mesh Class-AB architecture with minimum current selectors, the amplifier achieves high drive current capability under strict sub-1.0 V internal headroom constraint. The integrated low-dropout regulator reduces power supply sensitivity, achieving 50 dB PSRR at 100 kHz. In open-loop AC simulations, the operational amplifier demonstrates a DC gain of 147 dB, a UGBW of 59 MHz, and a phase margin of 68 degree with a 5-pF load. Applied within an 8 V/V switched capacitor gain stage, FFT spectral analysis shows an SNDR of 78.2 dB and an SFDR of 78.6 dBc while consuming 2.45 mW at a 104.2 kHz clock frequency and occupying 0.0946 mm^2 of silicon area. These results demonstrate that the proposed architecture is highly attractive for energy-efficient, high-precision mixed-signal SoCs in advanced FD-SOI nodes.

References

1. Mahattanakul et al., "Design procedures for three-stage CMOS op-amps with nested Miller compensation," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 52, no. 8, pp. 1508–1514, 2005
2. M. Yoshioka et al., "A 10b 50MS/s 8.2mW SAR ADC in 65nm CMOS," IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers, pp. 452–453, 2007.
3. B. R. Gregoire et al., "An oversampled ADC with dynamic element matching," IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers, pp. 540–541, 2008.
4. R. Hogervorst et al., "A compact power-efficient 3 V CMOS rail-to-rail input/output operational amplifier for a 3-V CMOS cell library," IEEE Journal of Solid-State Circuits, vol. 29, no. 12, pp. 1505–1513, 1994.
5. K. J. de Langen et al., "Compact low-voltage power-efficient operational amplifier cells," IEEE Journal of Solid-State Circuits, vol. 33, no. 10, pp. 1482–1496, 1998.

From Head Trajectory Diagrams to Deep Gait Recognition: A Single Head-Worn Sensor Approach to Gait Symmetry Visualization and Person Identification

Tong-Hun Hwang and Alfred O. Effenberg

Institute of Sports Science, Leibniz University Hannover, 30167 Hannover, Germany

tonghun.hwang@sportwiss.uni-hannover.de

Abstract. Gait is both a clinical marker of health and a unique biometric feature, yet conventional gait analysis and gait recognition systems remain complex and difficult to deploy outside laboratories. This paper consolidates a single head-worn inertial sensing approach that addresses both goals from one feature: the trajectory of the head during walking. Head trajectories of twelve healthy adults were captured with an inertial measurement unit (IMU) based motion-capture system on a 400-m running track. First, the trajectories were segmented by head vertical peaks, transformed into a walking-vector-aligned local coordinate system, and visualized through two gait eye diagrams and a gait W-diagram inspired by eye patterns in digital communications, enabling intuitive assessment of lateral and vertical gait symmetry; step counting reached a mean absolute percentage error below 0.65%. Second, the same three-dimensional head trajectories, segmented into 586 strides, were used to train a compact deep neural network that identifies individuals from a 64×3 input. The model reached an average identification accuracy of up to 94.4% and an equal error rate between 2.33% and 4.13% across practical scenarios. Together, these results show that head trajectory alone supports both gait symmetry visualization and reliable person identification, pointing toward head-worn devices for the diagnosis and recognition of pathological gait.

Keywords: Gait Symmetry, Gait Recognition, Head Trajectory, Head-Worn Sensor, Inertial Measurement Unit, Deep Learning.

1. Introduction

Human gait carries two complementary kinds of information. In the clinic, the symmetry of walking is a key indicator in the diagnosis and rehabilitation of pathological gait, supporting the management of neurological, muscular, and sensorimotor conditions such as post-operative hip and knee symptoms, concussion, stroke, and Parkinson's disease [1]. In security and forensics, gait is a distinctive biometric that remains useful even when the face is hidden or clothing has changed, and research on automatic gait recognition has grown rapidly with deep learning [2,3].

Both applications, however, have traditionally relied on elaborate measurement systems—camera arrays, instrumented floors, or full-body sensor suits—that are difficult to deploy in daily life and that limit the size of available datasets. At the same time, head-worn devices such as earbuds, smart glasses, and head-mounted displays have become common consumer products with onboard inertial sensors, offering a seamless and long-lived measurement platform [4,5]. This motivates a single question: how much of both the clinical and the biometric content of gait can be recovered from the trajectory of the head alone?

The head is a well-suited proxy for whole-body gait because it sways laterally and oscillates vertically in coordination with the trunk and pelvis during walking [6]. This paper brings together two studies built on this idea. The first develops visualization methods that transform head trajectory into readable symmetry diagrams; the second demonstrates that the same trajectory, processed by a compact deep learning model, is sufficient to identify individuals. Presenting them jointly clarifies a practical path from a single, easily worn sensor toward both the diagnosis and the recognition of gait.

2. Experiments and Discussion

2.1 Data Acquisition and Coordinate Transformation

The head movements of twelve healthy participants (six female, six male; age 29.8 ± 6.8 years; height 173.3 ± 8.6 cm) were recorded while they walked about one hundred steps on the straight section of a 400-m outdoor running track. Each participant wore seventeen IMUs of the Xsens MVN Awinda system, which reconstructs three-dimensional joint positions at 60 Hz. Only the head position was used for the analyses reported here. The walking path was segmented into steps at the automatically detected vertical peaks of the head. For each step, a walking vector defined between successive peak positions established a local coordinate

system in which the anterior axis aligns with the walking direction and the lateral axis points to the left, making trajectories from different absolute positions and directions directly comparable (Fig. 1).

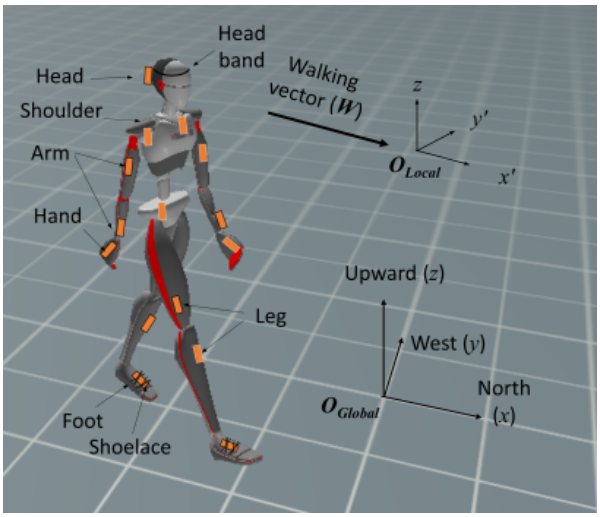


Fig. 1. Sensor placement of the IMU-based motion-capture suit and the coordinate systems: the global frame (O_{Global}; x north, y west, z up) and the walking-vector-aligned local frame (O_{Local}; x′, y′, z) used to align each stride. Reproduced from [2].

2.2 Gait Symmetry Visualization

Adopting the concept of the eye diagram from digital communications, the segmented head trajectories were overlaid to form three diagrams: two gait eye diagrams on the transverse plane (one preserving step length, one normalizing it) and a gait W-diagram on the frontal plane. The names describe the overlaid shapes: superimposed left and right steps open into an eye—just as repeated pulses do in the eye pattern engineers use to check signal quality—while on the frontal plane the head traces the letter W over one stride. Both are projections of the same three-dimensional aligned stride (Fig. 2c)—exactly the form later used as input to the recognition model. Left and right steps appear as the two lobes of each pattern, so that asymmetry in lateral range, vertical oscillation, and step alignment becomes visible at a glance. From these diagrams, parameters analogous to eye height, jitter, bandwidth, and signal-to-noise ratio were defined and summarized with three established gait symmetry indices (ratio index, symmetry index, and gait asymmetry). Because every parameter depends on detecting head events once per step, step-counting accuracy served as the validation criterion: the mean absolute percentage error was below 0.65% for total steps and at most 2.13% for left–right step recognition, confirming that a single head-worn sensor can independently characterize gait symmetry (Fig. 2) [1].

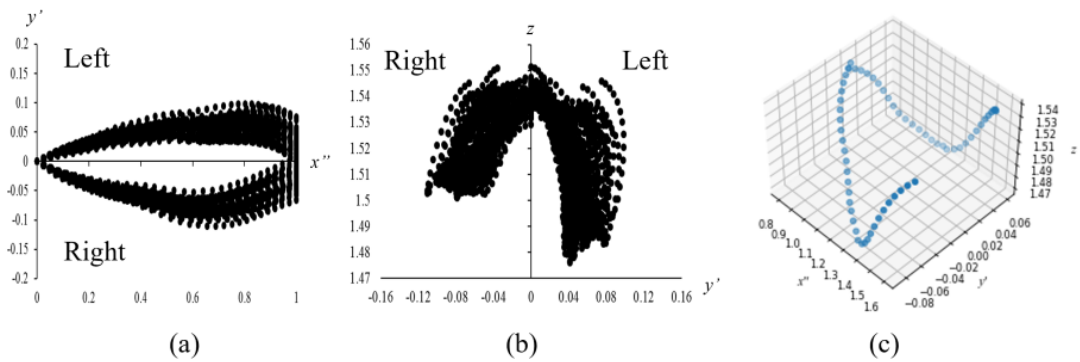


Fig. 2. Head trajectory of a walker aligned by the walking vector, shown as (a) a gait eye diagram (normalized) on the transverse plane (x″–y′), (b) a gait W-diagram on the frontal plane (y′–z), and (c) the full three-dimensional aligned stride from which both projections are taken. Left and right steps form the two lobes; differences between them reveal lateral and vertical gait asymmetry. Panels (a) and (b) are adapted from [1]; (c) is newly rendered from the same head-trajectory data.

2.3 Deep Gait Recognition from Head Trajectory

The same three-dimensional head trajectories were then repurposed for person identification. A pair of left and right steps (one stride) formed a single input, yielding 586 strides across the twelve participants. A compact neural network—comprising normalization, flattening, and three dense layers with rectified-linear activation, trained with sparse categorical cross-entropy—received an input of only 64 sampled three-dimensional positions (64×3) and output one of twelve identities (Fig. 3). To reflect realistic conditions, three covariates were varied: sampling rate (5–60 Hz), the number of training strides, and the inclusion of body height. Each condition was repeated ten times with different training and validation splits to mitigate bias from the small sample.

Including body height, the model reached an average identification accuracy of 94.4% when trained on thirty strides at 60 Hz, with an equal error rate of 2.33% and an area under the one-versus-rest receiver operating characteristic (ROC) curve close to 1.0. Performance remained practical under degraded conditions: 90.1% accuracy with only five strides at 20 Hz including height, and 86.5% with ten strides at 20 Hz when height was excluded, the latter giving an equal error rate of 4.13%. Notably, the 64×3 input is far smaller than the 88×128 images used in silhouette-based recognition or the 32×32 maps used in ground-reaction-force methods, lowering computational and storage costs [2,5].

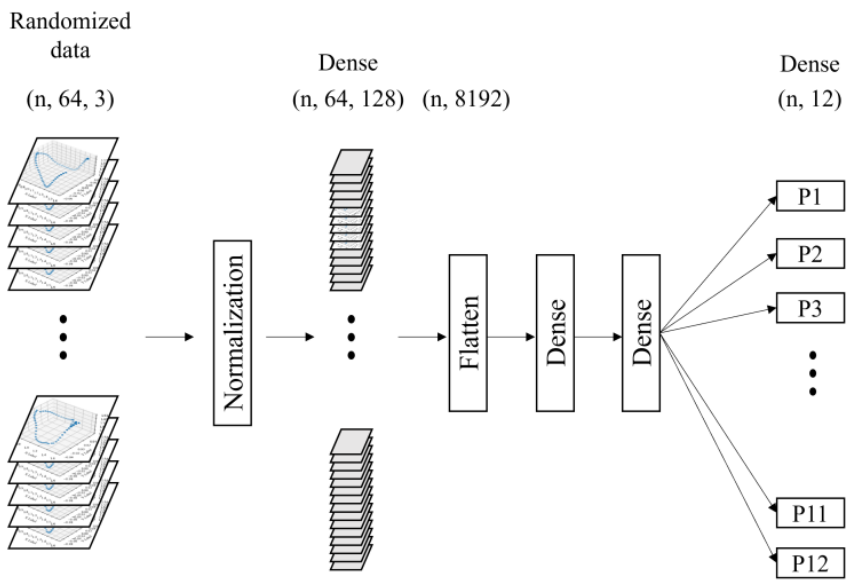


Fig. 3. Architecture of the deep learning model: a normalization layer, a flatten layer, and three dense layers map each 64×3 head-trajectory stride to one of twelve participant classes (P1–P12). Adapted from [2].

2.4 Discussion

The two results are complementary. That 94.4% accuracy is reached from an input as small as 64×3 shows that a single stride already carries a strong individual signature, consistent with the earlier observation that diagram shapes are individually distinct [1]. The same head sway that encodes identity also encodes left–right symmetry, so one lightweight feature serves a biometric and a clinical purpose at once. Clinically, interpretability matters as much as accuracy: asymmetry appears directly as a difference between the two lobes, giving the low-barrier feedback that wearable gait monitoring has long sought [7], and the intuitive diagrams could also drive gait sonification for rehabilitation [8]. Because head trajectory can be recovered from a camera or an IMU alike, clinic- and home-based recordings share a common structure and can be compared. Cohort homogeneity also conditions the identification result: all twelve participants were healthy adults, so the network separated individual signatures within the narrow range of healthy gait. Markedly different patterns—neurological, post-stroke, or post-operative—would be expected to be more separable, yet raise the question of whether a deviation marks an individual or a pathology, and whether identification stays stable as gait changes during rehabilitation; answering this requires clinical data. Recovering reliable head position from a single IMU is a further limitation that complementary sensing could mitigate.

3. Conclusions

Head trajectory alone, captured by a single head-worn sensor, proved sufficient both to visualize gait symmetry and to identify individuals—showing that one simple feature can carry clinical and biometric information otherwise obtained from far more elaborate systems. The next steps are to run the eye and W-diagrams in real time on consumer head-worn devices such as earbuds and head-mounted displays, with visual and acoustic feedback, and to condense the diagram parameters into a single interpretable asymmetry index tracked over time and compared across sessions and devices. Above all, the approach should be validated on clinical populations—Parkinson's disease, post-stroke, or post-concussion gait—and on larger cohorts, so that it can support the diagnosis and rehabilitation of pathological gait beyond the laboratory.

References

1. Hwang, T.-H., Effenberg, A.O.: Head trajectory diagrams for gait symmetry analysis using a single head-worn IMU. Sensors 21(19), 6621 (2021).
2. Hwang, T.-H., Effenberg, A.O.: Simple head trajectory measurement for deep gait recognition. IEEE Sensors J. 24(22), 37144–37151 (2024).
3. Sepas-Moghaddam, A., Etemad, A.: Deep gait recognition: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 45(1), 264–284 (2023).
4. Shen, Y., et al.: GaitLock: Protect virtual and augmented reality headsets using gait. IEEE Trans. Dependable Secure Comput. 16(3), 484–497 (2019).
5. Hwang, T.-H., Reh, J., Effenberg, A.O., Blume, H.: Real-time gait analysis using a single head-worn inertial measurement unit. IEEE Trans. Consum. Electron. 64(2), 240–248 (2018).
6. Kavanagh, J.J., Morrison, S., Barrett, R.S.: Coordination of head and trunk accelerations during walking. Eur. J. Appl. Physiol. 94, 468–475 (2005).
7. Shull, P.B., Jirattigalachote, W., Hunt, M.A., Cutkosky, M.R., Delp, S.L.: Quantified self and human movement: A review on the clinical impact of wearable sensing and feedback for gait analysis and intervention. Gait Posture 40(1), 11–19 (2014).
8. Effenberg, A.O., Fehse, U., Schmitz, G., Krueger, B., Mechling, H.: Movement sonification: Effects on motor learning beyond rhythmic adjustments. Front. Neurosci. 10, 219 (2016).

From Sinusoids to Chirps: Rethinking the Wireless Physical Layer for 6G and Beyond

Hyeon Seok Rou

School of Computer Science and Engineering, Constructor University,
Campus Ring 1, 28759 Bremen, Germany
hsrou@ieee.org

Abstract. Orthogonal frequency-division multiplexing (OFDM) and its static sinusoidal subcarriers have underpinned the fourth- and fifth-generation (4G/5G) wireless systems for nearly two decades, delivering high spectral efficiency and resilience to multipath fading through an efficient multicarrier architecture. However, as sixth-generation (6G) systems move toward high-mobility applications such as vehicular networks, unmanned aerial vehicles, and low Earth orbit satellite constellations, operating at ever-higher frequency bands, the wireless channel becomes doubly dispersive, and the time-invariance assumption underlying OFDM no longer holds, with Doppler-induced degradation becoming prominent at the waveform level. In this article, we revisit the evolution of the wireless physical layer through the lens of a pendulum-like oscillation between two competing design priorities, rate and robustness, and argue that the pendulum is now swinging back toward robustness, calling for a resurrection of spectrum spreading - this time jointly over time and frequency, specifically for Doppler resilience. Among the emerging candidate waveforms, we highlight affine frequency division multiplexing (AFDM), whose parameterizable chirp subcarriers generalize the sinusoids of OFDM, achieving full diversity in doubly dispersive channels and native suitability for integrated sensing and communications, while retaining maximal backward compatibility with the mature OFDM ecosystem. Overall, we argue that the transition from sinusoids to chirps is a technically motivated and compelling evolutionary direction for the future wireless physical layer.

Keywords: 6G, Physical Layer, Wireless Communications, Waveform Design, Doubly Dispersive Channels, Affine Frequency Division Multiplexing (AFDM), Chirps, Integrated Sensing and Communications (ISAC).

1. The Rate-Robustness Pendulum of Wireless Evolution

The evolution of wireless cellular systems is often summarized as a monotonic trajectory toward ever-higher data rates, with the gradual inclusion of new applications and features: from analog voice connectivity in the first generation (1G), to digital traffic in 2G, mobile internet in 3G, broadband connectivity in 4G, and the service flexibility of 5G, which supports enhanced mobile broadband, massive machine-type communications, and ultra-reliable low-latency communications [1]. More recently, sixth-generation (6G) systems have been envisioned around AI-native operation and integrated sensing and communications (ISAC) [2].

A closer examination, however, reveals that this history is more accurately represented not as a one-dimensional progression along the rate axis, but as a three-dimensional trajectory in the joint *rate-robustness-time* space, as illustrated in Fig. 1, in which the conventional rate-centric narrative is recovered merely as a projection of the true, richer trend - a projection that obscures the oscillating design priority along the robustness axis. Namely, 2G pursued robustness against transmission errors through digitalization and channel coding; 3G pursued robustness against multiuser interference through wideband spectrum spreading [3]; 4G swung back toward rate, prioritizing mobile broadband through narrowband multiplexing over the sinusoidal subcarriers of orthogonal frequency-division multiplexing (OFDM) [4,5]; and 5G further enhanced rate and flexibility through massive multiple-input multiple-output (MIMO) beamforming and advanced signal processing [6]. In particular, the 3G era anchored its architecture in code-division multiple access (CDMA), where each user's narrowband signal is spread across the full bandwidth by a unique pseudo-noise chip sequence, with the wideband occupancy also providing inherent frequency diversity as an intrinsic waveform property.

To motivate the architectural realignment ahead, this evolution can be cast as a *dialectic rhythm* composed of four stages [7]: the *thesis* - the era of interference robustness through spectrum spreading (3G); the *antithesis* - the era of rate and efficiency through multiplexing and MIMO (4G/5G); the *crisis* - the Doppler bottleneck arising from high mobility and high carrier frequencies (6G); and the *synthesis* - the era of Doppler robustness

through the reintroduction of spectrum spreading (6G and beyond).

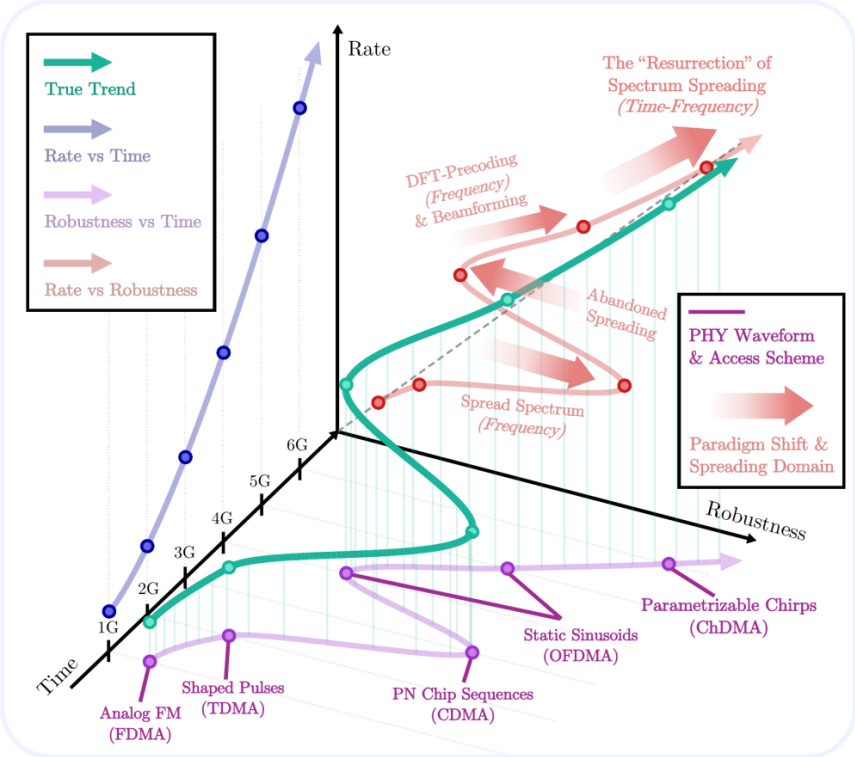


Fig. 1. The wireless physical-layer evolution represented in the three-dimensional rate-robustness-time space, revealing the pendulum-like oscillation of design priorities between the two key engineering philosophies of rate and robustness across the generations, with the conventional rate-centric view recovered as a projection of this richer underlying trend. The labels indicate the elementary per-symbol waveform and multiple-access scheme of each era: analog frequency modulation (FDMA), shaped time-domain pulses (TDMA), pseudo-noise chip sequences (CDMA), static sinusoids (OFDMA), and parametrizable chirps (ChDMA). Adapted from [Ref. 7, Fig. 1].

2. The Doppler Crisis of 6G

The success of OFDM in 4G and 5G rests on the fact that its static sinusoidal subcarriers are eigenfunctions of linear time-invariant channels, enabling elegant per-subcarrier equalization and resilience to multipath fading with a simple cyclic prefix (a short repetition of the symbol tail that renders the channel circular and easily equalizable), all implemented via the computationally efficient fast Fourier transform (FFT) [4]. This time-invariance condition was generally satisfied in the low-mobility, sub-6 GHz operating regimes of 4G/5G, but becomes increasingly difficult to maintain in many of the envisioned 6G use cases: in high-mobility scenarios such as vehicle-to-everything (V2X) networks, unmanned aerial vehicles (UAVs), and low Earth orbit (LEO) satellite constellations, the wireless channel becomes *doubly dispersive*, varying rapidly in both time and frequency. This is further exacerbated by the planned migration toward higher frequency bands including the upper mid-band, millimeter-wave, and sub-terahertz spectrum, where even moderate-to-high velocities induce significant Doppler shifts (e.g., 120 km/h at a 71 GHz carrier corresponds to approximately 7% of a typical 120 kHz subcarrier spacing), thereby impairing subcarrier orthogonality and causing detrimental inter-carrier interference (ICI).

Mitigation strategies do exist within the OFDM framework, but they require additional overhead and complexity, or are suitable only for specific operating regimes; Doppler pre-compensation, for instance, is effective for predictable line-of-sight-dominated motion such as single-LEO orbital trajectories, but becomes challenging under unpredictable mobility with rich multipath scattering. More fundamentally, such approaches only *map* coded information across non-spreading subcarriers rather than achieving intrinsic waveform-level *spreading*. This coded diversity becomes less effective for narrowband allocations - control signals, synchronization channels, and most importantly the small data packets that dominate practical traffic [8] - which cannot span sufficiently distinct channel realizations, and the waveform itself must therefore carry the required robustness. Moreover, the emerging ISAC paradigm reframes the concept of Doppler entirely [9]: while communications has so far treated it as an impairment to be mitigated or compensated, sensing *relies* on

it as a primary observable quantity for estimating the velocity of environmental scatterers and targets, and a Doppler-fragile waveform is therefore ill-suited to the resource-efficient, multifunctional vision of 6G and beyond.

3. The Resurrection of Spectrum Spreading

A promising waveform direction is therefore to employ per-symbol waveforms that inherently occupy an extended region of the time-frequency (TF) plane, distributing each symbol's energy across *both* time and frequency as an intrinsic structural property. In other words, waveform design must shift its primary focus back to robustness, and a form of spectrum spreading must return to the physical layer - but this time not for multiuser interference suppression as in 3G, nor for frequency diversity, which coded wideband systems already achieve, but specifically for *Doppler resilience*. At the same time, ever-increasing rate demands remain a key priority, precluding a naive return to CDMA-style frequency-domain wideband spreading; instead, the 6G paradigm calls for *joint time-frequency spreading*, while preserving the multiplexing architecture, spectral efficiency, and ecosystem maturity of the OFDM legacy.

Two fundamentally different philosophies have emerged to realize this joint TF spreading. The first is *2D isotropic spreading* via the delay-Doppler (DD) domain, exemplified by the orthogonal time frequency space (OTFS) waveform family [10] and its descendants such as Zak-OTFS [11] and orthogonal delay-Doppler division multiplexing [12], in which information symbols are placed on a 2D DD grid and spread across the entire allocated TF plane via a 2D transform. The DD-domain construction offers key structural advantages: a doubly dispersive channel that fluctuates rapidly in the TF domain maps to a nearly time-invariant, sparse, point-like representation in the DD domain, and the DD coordinates translate directly into the range and radial velocity of scatterers, making the paradigm intrinsically sensing-friendly and a natural fit for ISAC [13].

These strengths, however, come with practical design considerations. As a consequence of the Balian-Low theorem, no pulse can be simultaneously well-localized in both time and frequency at the critical sampling density, and achieving well-behaved DD-domain localization therefore requires dedicated pulse-shaping machinery with associated time and bandwidth expansion overhead. In addition, DD-grid-tailored embedded pilots (known reference symbols used for channel estimation) incur higher overhead than conventional OFDM pilot scattering (a gap that recent superimposed- and spread-pilot schemes have partly addressed), and the resource granularity becomes the full 2D DD grid rather than the 1D subcarrier, introducing a nontrivial mapping between the DD grid and the existing cellular numerology, i.e., the standardized set of subcarrier spacings and frame parameters, for multiuser multiplexing and scheduling.

These considerations do not diminish the technical maturity of the 2D isotropic paradigm, but they do motivate a natural question: can the structural benefits of joint TF spreading - a sparse effective channel, full TF diversity, and intrinsic Doppler resilience - be obtained through a construction that stays closer to the OFDM legacy? The second philosophy, *sheared spreading* via parameterizable chirps, exemplified by affine frequency division multiplexing (AFDM) [14], answers this question affirmatively.

4. From Sinusoids to Chirps: AFDM

A linear chirp is a signal whose instantaneous frequency varies linearly in time, parameterized by the chirp rate, such that the sinusoid is recovered as the special case when the chirp rate is zero, as illustrated in Fig. 2a. The relationship between the sinusoid and the chirp is therefore one of *generalization* rather than substitution: introducing a chirp rate adds a new degree of freedom without discarding any property of the sinusoidal baseline. Importantly, chirps are not new to wireless systems: they have long been the foundation of frequency-modulated continuous-wave (FMCW) radar, underpin the long-range low-power connectivity of LoRa, and form the subcarrier set of orthogonal chirp division multiplexing (OCDM) [15].

The fundamental motivation for the shift is channel-theoretic, and is best appreciated geometrically on the TF plane, as illustrated in Fig. 2b and Fig. 2c. For a sinusoid, delay and Doppler act on *different* signal properties: a delay produces only a phase rotation that leaves the waveform unchanged, while a Doppler shift translates the signal to a fundamentally different sinusoid. For a chirp, by contrast, both delay and Doppler act on the *same* property, namely the starting frequency, yielding a single net displacement while the chirp rate and waveform shape are fully preserved. The structural significance of this behavior is the collapse of the two-dimensional delay-Doppler effect into a single scalar displacement over a preserved waveform family, with the chirp rate acting as a tunable coupling coefficient between delay and Doppler. Indeed, by selecting the chirp rate according to the full-diversity condition [14], which ensures that each symbol benefits from all

independent propagation paths, the net displacements of all distinct paths of a doubly dispersive channel are guaranteed to be mutually distinct in the effective channel. This converts the dense, unstructured ICI experienced by sinusoidal subcarriers into a well-organized, sparse, quasi-diagonal structure, in which each received symbol overlaps with only a few, predictable neighbors, enabling efficient channel estimation and equalization.

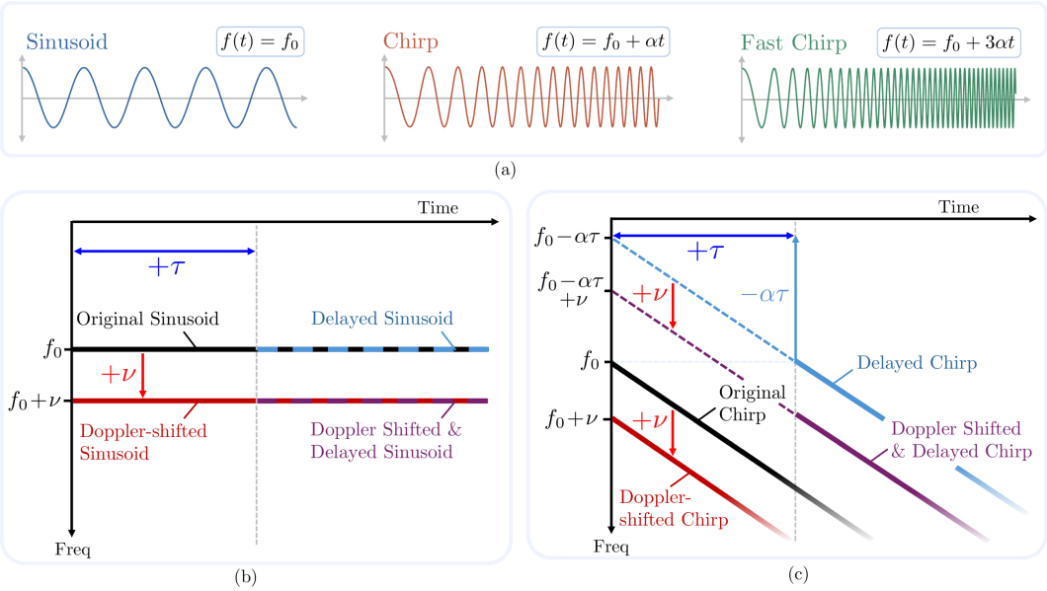


Fig. 2. Illustration of sinusoid and chirp behaviors: (a) time-domain waveforms, highlighting the sinusoid as the special case of zero chirp rate; (b) instantaneous frequency of sinusoids under delay and Doppler, where a delay preserves the frequency while a Doppler shift maps the subcarrier to a distinct sinusoid; and (c) instantaneous frequency of chirps, where both effects act on the starting frequency alone, yielding a net displacement while the chirp rate and the waveform shape are fully preserved. Adapted from [Ref. 7, Fig. 3].

Following this principle, AFDM builds a multicarrier waveform from chirp subcarriers via the discrete affine Fourier transform (DAFT), a generalized FFT parameterized by two chirp rates, customarily denoted c_1 and c_2 [14]. The setting $(c_1, c_2) = (0, 0)$ recovers the FFT of OFDM, while $(1/2N, 1/2N)$, with N denoting the number of subcarriers, yields OCDM, and AFDM therefore includes both as special cases within a single unified framework in which transitioning between waveform modes requires only a parameter adjustment in response to channel conditions and application demands [16]. The two parameters play asymmetric roles: c_1 controls the slope of the chirps in the TF plane, and hence the extent of TF spreading and the aforementioned diversity guarantee. The second parameter c_2, on the other hand, does not affect subcarrier orthogonality or the diversity condition at all, and instead acts as a free precoding degree of freedom, which has been exploited for shaping the waveform's delay-Doppler radar response (the ambiguity function), peak-to-average power ratio reduction, index modulation, and physical-layer security - a flexibility that fixed waveforms such as OFDM or OTFS structurally lack.

It should be highlighted that this generalization comes at a low migration cost [17]. The inverse DAFT decomposes exactly into a sample-wise quadratic-phase multiplication, a standard inverse FFT, and a second sample-wise quadratic-phase multiplication; therefore, any FFT-based baseband designed for OFDM can be extended to AFDM by inserting two elementwise, unit-magnitude phase rotations around the existing core, adding an $O(N)$ overhead on top of the unchanged $O(N \log N)$ FFT, while the analog front-end, radio-frequency handling, and existing numerology remain intact. The subcarriers themselves remain the natural multiple-access resource, allowing legacy orthogonal frequency-division multiple access logic to carry over directly to its affine counterpart. Moreover, the inherent chirp structure renders AFDM natively suitable for ISAC: echo signals can be processed using mature FMCW-like radar principles, and the delays and Doppler shifts of targets can be estimated jointly with communication channel estimation from a single frame [13,18,19].

5. The Road Ahead

Since the formal introduction of AFDM in 2021 and its consolidation in [14], a broad and still-expanding

research landscape has emerged across nearly every axis of physical-layer design, spanning mathematical foundations, transceiver signal processing, MIMO and intelligent-surface integration, multiple access, and ISAC, in which the parameterizable chirp structure has demonstrated applicability well beyond Doppler-resilient communications [20]. Some of the most important remaining challenges are integrative rather than conceptual. On the standardization side, per-resource-block chirp-parameter signaling must be incorporated into the 6G numerology, with relevant discussions having recently begun within the 3rd Generation Partnership Project (3GPP), the body that standardizes cellular radio interfaces. The small and interpretable chirp-parameter space is also an unusually tractable target for AI-native waveform adaptation, far more amenable to learning-based approaches than end-to-end waveform design. Finally, over-the-air testbeds and field trials integrated with standardized numerology remain comparatively scarce, and their development will be essential to inform standardization decisions.

In all, the transition *from sinusoids to chirps* should be viewed not as a rejection of the OFDM ecosystem, but as a natural and technically motivated evolution and extension of it. Rather than a terminal design choice, the parameterizable chirp framework is an entry point into a uniquely flexible design space that the community has only begun to explore. As 6G and beyond move toward high mobility, high-frequency operation, and multifunctional wireless services, we advocate that chirp-based waveforms, best exemplified by AFDM, represent a promising direction for a robust, flexible, and implementation-conscious physical layer.

References

1. ITU-R, “IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond,” Recommendation ITU-R M.2083-0, Geneva, Switzerland, 2015.
2. R. Liu, L. Zhang, R. Y.-N. Li, and M. Di Renzo, “The ITU vision and framework for 6G: Scenarios, capabilities, and enablers,” IEEE Veh. Technol. Mag., vol. 20, no. 2, pp. 114–122, 2025.
3. A. J. Viterbi, CDMA: Principles of Spread Spectrum Communication. Reading, MA, USA: Addison-Wesley, 1995.
4. R. van Nee and R. Prasad, OFDM for Wireless Multimedia Communications. Boston, MA, USA: Artech House, 2000.
5. E. Dahlman, S. Parkvall, and J. Sköld, 4G: LTE/LTE-Advanced for Mobile Broadband. Oxford, U.K.: Academic Press, 2013.
6. E. Björnson, E. G. Larsson, and T. L. Marzetta, “Massive MIMO: Ten myths and one critical question,” IEEE Commun. Mag., vol. 54, no. 2, pp. 114–123, 2016.
7. H. S. Rou, G. T. F. de Abreu, E. Björnson, S. Kim, and M. Kountouris, “The resurrection of spectrum spreading for 6G and beyond: From sinusoids to chirps,” arXiv preprint arXiv:2605.00249, 2026.
8. M. Shafi et al., “Industrial viewpoints on RAN technologies for 6G,” arXiv preprint arXiv:2508.08225, 2025.
9. F. Liu et al., “Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond,” IEEE J. Sel. Areas Commun., vol. 40, no. 6, pp. 1728–1767, 2022.
10. R. Hadani et al., “Orthogonal time frequency space modulation,” in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), 2017, pp. 1–6.
11. S. K. Mohammed, R. Hadani, A. Chockalingam, and R. Calderbank, “OTFS - Predictability in the delay-Doppler domain and its value to communication and radar sensing,” IEEE BITS Inf. Theory Mag., vol. 3, no. 2, pp. 7–31, 2023.
12. H. Lin and J. Yuan, “Orthogonal delay-Doppler division multiplexing modulation,” IEEE Trans. Wireless Commun., vol. 21, no. 12, pp. 11024–11037, 2022.
13. H. S. Rou et al., “From orthogonal time-frequency space to affine frequency-division multiplexing: A comparative study of next-generation waveforms for integrated sensing and communications in doubly dispersive channels,” IEEE Signal Process. Mag., vol. 41, no. 5, pp. 71–86, 2024.
14. A. Bemani, N. Ksairi, and M. Kountouris, “Affine frequency division multiplexing for next generation wireless communications,” IEEE Trans. Wireless Commun., vol. 22, no. 11, pp. 8214–8229, 2023.
15. X. Ouyang and J. Zhao, “Orthogonal chirp division multiplexing,” IEEE Trans. Commun., vol. 64, no. 9, pp. 3946–3957, 2016.
16. H. Yin et al., “Affine frequency division multiplexing: Extending OFDM for scenario-flexibility and resilience,” IEEE Wireless Commun., vol. 32, no. 6, pp. 200–208, 2025.
17. H. S. Rou and G. T. F. de Abreu, “AFDM as a software upgrade of OFDM: One firmware patch, a new frontier,” arXiv preprint arXiv:2605.23062, 2026.
18. A. Bemani, N. Ksairi, and M. Kountouris, “Integrated sensing and communications with affine frequency division multiplexing,” IEEE Wireless Commun. Lett., vol. 13, no. 5, pp. 1255–1259, 2024.
19. K. R. R. Ranasinghe, H. S. Rou, G. T. F. de Abreu, T. Takahashi, and K. Ito, “Joint channel, data, and radar parameter estimation for AFDM systems in doubly-dispersive channels,” IEEE Trans. Wireless Commun., vol. 24, no. 2, pp. 1602–1619, 2025.
20. H. S. Rou, K. R. R. Ranasinghe, V. Savaux, G. T. F. de Abreu, D. González G., and C. Masouros, “Affine frequency division multiplexing (AFDM) for 6G: Properties, features, and challenges,” IEEE Commun. Stand. Mag., 2026.

Unearthing Silent Data with Process Design

Jae Sook Cheong

University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany
jaesookc@gmail.com

Abstract. We live in technology-driven societies heavily shaped by Artificial Intelligence. While modern AI performance depends heavily on data quality, data bias remains a critical hurdle. Among various sources of bias, an often-overlooked category is silent data, which typically stems from organizations with low digital maturity. These institutions struggle with foundational data collection and quality maintenance, ultimately hindering their strategic operational goals. To address this systemic challenge, this paper proposes comprehensive protocols for managing complex research projects and related administrative operations. Our methodological approach centers on structured data collection into two central databases paired with systematic quality enhancement. Furthermore, we introduce a resource-efficient data quality management strategy specifically designed to handle diverse, scattered information sources under resource constraints. Presented as a case study, this work demonstrates how to optimize data intake and maintenance within the data infrastructure in academic and research environments. Ultimately, our approach helps bring silent data to light, actively mitigating data bias in AI models.

Keywords: processes for digitalization, data quality, bias, silent data

1. Introduction

We live in technology-driven societies, especially shaped by Artificial Intelligence (AI). The performance of AI models is a universal concern, as it heavily depends on the quality of training data. Data bias has gained considerable attention, prompting extensive research on the topic. Data bias can stem from representation, history, measurement, aggregation, evaluation, and other factors. However, little attention has been given to silent data, which contributes to representation bias.

Silent Data refers to data that are not collected at all or remain unincorporated into digital systems, effectively rendering them invisible within an organization. Such data typically emerge in environments with low levels of digital maturity or unstructured administrative processes and lead to representation bias that distorts the training base of AI models. It often comes from under-digitized organizations, which is a significant portion of all the organizations. An estimated 90 percent of organizations are undergoing some form of digital transformation according to McKinsey research [1], though this primarily applies to companies of a certain size. Smaller firms that lack the resources for comprehensive digitalization, increasing their risk of falling behind. As illustrated in Figure 1, companies can be categorized into four levels of digital maturity [2]: Beginners, Conservatives, Fashionistas, and Digirati (ranked from low to high). This report [2] published in 2014 indicated that 65 percent of companies was at the Beginners-level.

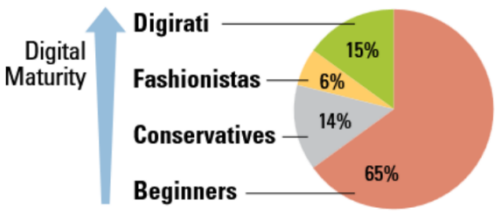


Fig. 1. The ratio of companies in digital maturity levels.

Organizations function as a machinery that enable people to achieve ambitious, otherwise unattainable goals. Organizations operate typically through hierarchical structures, where decisions are made, communicated, interpreted at each group, and ultimately implemented by staff. This requires strategic planning, well-defined

procedures, transparent and effective communication channels, and regular progress evaluation. Effective organizational performance requires careful balancing of human dynamics and structural boundaries to achieve complex objectives. The inherent diversity of personalities and roles within any organization naturally introduces uncertainty in collaborative efforts. An appropriate level of processes and rules with clearly defined responsibilities will help the organization with making smoother collaborations and protecting the people by mitigating internal and external risks. While organizations evolve organically, accumulating diverse stakeholders, practices, and historical baggage, digitalization on administrative operations demands coherent and systematic rules and standardized data. Creating efficient digital systems that harmonize this diversity presents significant challenges.

Such diversities are exemplified by the Africa Multiple project with its multicultural membership spanning varied languages, ethnicities, and traditions. Colonial legacies further complicate matters, evident in persistent inconsistencies like geographical naming conventions. In this context, the project faces two challenges: (1) data collection from diverse and scattered sources across academic research projects and in administrative groups, and (2) aligning organizational processes to enable seamless data flow and information sharing. This paper proposes a strategic process for data collection and maintenance across two primary databases, establishing optimized workflows to support both immediate project needs and long-term digital transformation.

2. Our Approach

Our approach begins with establishing reliable data and information sources aligned with project objectives. These sources serve dual purposes: facilitating transparent communication among project members and staffs, and effectively showcasing the project outcome and progress to the external stakeholders. To design a well organized and effective work processes and dataflow, we take three key criteria: data reliability, data consistency across databases, and process simplicity. Based on these criteria, we implemented structured procedures for robust data collection and data quality assurance with the following steps:

• design databases with clearly defined purposes.

• identify all required data collection procedures.

• develop data collection process for each data category.

• categorize data into Focus Groups to maintain data quality with limited resources.

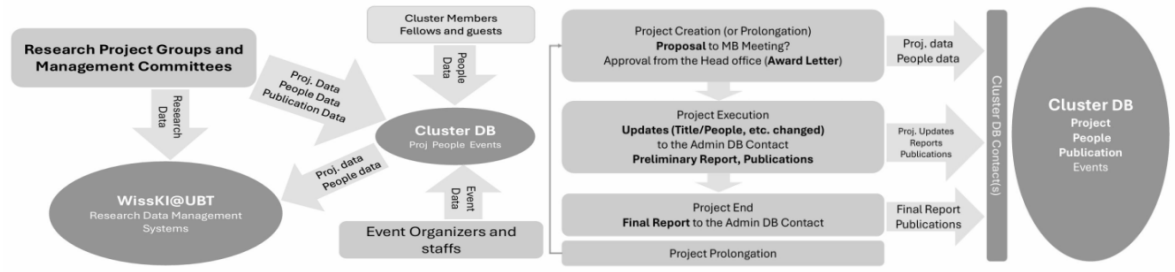
3. Data Collection and Maintaining

The main outcome of the Africa Multiple project is research activities in African Studies for reconfiguring African studies through interdisciplinary collaborations. This can be optimally supported by an appropriate research data management system. Integrated with multiple Research Data Management System (RDMSs), WissKI@UBT serves as the primary research data management system. This showcases the main output of the project–research data in African Studies. The second one is the Cluster DB, which maintains operational data concerning personnel, projects, publications, and events. This administrative database facilitates the Cluster's daily operations while supporting its research activities and related administrative activities. These two systems share two common data categories. The first one is research project information, more precisely, project ID, project title, and duration–and the profiles of the relevant researchers of research groups. These datasets are drawn from the Cluster DB and fed into WissKI@UBT. To maintain these two databases reliable, we evaluate data quality using established metrics. Among all the data quality metrics presented in [3], we prioritize five fundamental criteria: completeness (ensuring all required data is present), uniqueness (eliminating duplicate entries), validity (verifying data conforms to required formats), timeliness (maintaining current information), and consistency (ensuring uniform data across systems).

Figure 2 (a) illustrates main dataflow among WissKI@UBT, the Cluster DB, and other related groups of people involved in the Cluster activities. This visualizes how data enter the administrative Cluster DB, and how research project information and their related people data are transferred from the Cluster DB to WissKI@UBT. The Cluster DB serves as the central repository for administrative data essential to supporting research activities within the Cluster. It systematically organizes information on personnel (people), projects, publications, and events. Figure 2 (b) outlines this structured workflow, demonstrating how project data is systematically collected and maintained throughout the research life cycle to ensure data integrity.

We implement distinct approaches for maintaining data quality for project, publication, people and event data. As explained above, research projects follow clearly defined life cycles from initiation to completion, enabling systematic data collection and ensuring high-quality project information in the Cluster DB. As an outcome of the research projects, it is also relatively straightforward to collect publication data. People and event data, however, present greater challenges due to their scattered sources across time and various contact points. This decentralized structure requires significant time and resources to maintain high data quality. To address this challenge, we employ a tiered strategy for people and event data managements, classifying entities into three prioritized Focus Groups. Figure 3 shows Focus groups for people data. Each group is assigned a corresponding level of resource allocation, as detailed below.

- Focus Group 1 comprises group of people that makes decisions and runs the Africa Multiple in the Cluster This group receives the highest priority for data accuracy and ongoing maintenance.
- Focus Group 2 includes project participants across the Cluster and partner institutions, as well as fellows. Information for this group is accurate upon initial creation and updated periodically.
- Focus Group 3 covers invited speakers, performers, and guests for Cluster events. Information for this group is updated infrequently.



(a) (b)

Fig. 2. (a) Dataflow among the groups of the Africa Multiple project, WissKI@UBT and the Cluster DB. (b) Dataflow for project data collection.

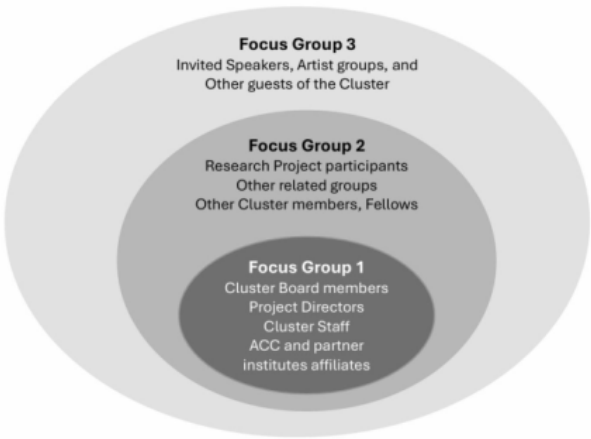


Fig. 3. Focus groups for the people data in the Cluster DB.

4. Conclusions

In this work, we proposed processes with optimal dataflow around two databases in the Cluster for data collection and for improving data quality. We also developed a strategy to effectively maintain the quality of the data collected from diverse and dispersed sources with limited resources. Initial implementation of these proposed processes has already yielded measurable benefits, especially in enhancing data quality and building a comprehensive quality assessment framework for project execution. It also has contributed to advancing digital and data literacy in the Cluster.

References

1. McKinsey &Company, Our Insight page, https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-digital-transformation, last updated 2024/8, first entered 2023/6.
2. Fitzgerald, M., Kruschwitz, F., Bonnet, D., Welch. M.: Embracing digital technology: A new strategic imperative. MIT sloan management review 55, no. 2 (2014)
3. Wikipedia Data Quality page, https://en.wikipedia.org/wiki/Data quality, last edited 2026/7/15.

Network-Enabled Distributed Physical AI: A 5G-Advanced and 6G Architecture for System 0–3 Robotic Intelligence

Harrison Hongseo Yun

Deutsche Telekom Innovation Laboratories, Winterfeldtstraße 21, 10781 Berlin
Harrison-hongseo.yun@telekom.de

Abstract. Physical Artificial Intelligence (Physical AI) is rapidly evolving from monolithic, device-centric robotic intelligence toward hierarchical architectures in which sensing, reactive control, reasoning, world modeling, and long-term learning operate at different temporal and computational scales. Emerging robotic AI architectures can be conceptually organized into four functional layers: System 0, responsible for high-frequency sensing, reflexes, and low-level control; System 1, responsible for fast learned perception-action policies and short-horizon behaviors; System 2, responsible for semantic reasoning, planning, and complex task execution; and System 3, responsible for world modeling, simulation, long-term learning, and fleet-level intelligence. These layers have fundamentally different requirements for response latency, communication bandwidth, computational capacity, reliability, and energy consumption. Consequently, executing the entire Physical AI stack locally on a robot is neither necessary nor, for increasingly capable foundation models, economically and energetically optimal. This white paper proposes a Network-Enabled Distributed Physical AI Architecture in which Physical AI functions are dynamically distributed across the robot, network edge, regional edge, and central AI cloud. 5G-Advanced and future 6G infrastructure becomes integral components of the robotic computing architecture rather than merely providing connectivity. The proposed architecture introduces differentiated communication and computing requirements for System 0–3, a hierarchical robot–edge–cloud computing framework, and network-aware AI orchestration capable of dynamically placing and migrating Physical AI workloads according to latency, bandwidth, reliability, compute availability, energy consumption, and operational context. The analysis indicates that System 0 should remain predominantly on-device because of hard real-time safety requirements, while System 1 is primarily local but may selectively exploit edge acceleration. System 2 represents the principal opportunity for 5G/6G edge computing, with target interaction latency on the order of tens to hundreds of milliseconds depending on the robotic task. System 3 can exploit regional and central AI infrastructure for large world models, digital twins, fleet learning, and simulation. This architecture establishes a pathway toward AI-native robotic networks, in which communications, computing, sensing, and Physical AI are jointly orchestrated as a distributed intelligence platform.

Keywords: 5G-A,6G, Network Capability Exposure, Distributed physical AI, Split computing,

1. Introduction

Artificial intelligence is moving beyond information processing in digital environments toward systems capable of perceiving, understanding, reasoning about, and interacting with the physical world.

Robots employing conventional control architectures have historically relied on deterministic pipelines consisting of perception, localization, planning, and control modules. Recent advances in robot foundation models are changing this architecture. Vision-Language Models (VLMs), Vision-Language-Action (VLA) models, generative action models, and World Models increasingly provide robots with generalized perception, semantic reasoning, task planning, and learned manipulation capabilities.

However, increasing model capability introduces an important systems problem.

A robot has strict physical constraints on:

- computational capability;
- memory capacity
- electrical power

- battery capacity
- thermal dissipation
- physical size and weight
- hardware cost

At the same time, the computational requirements of advanced Physical AI models are rapidly increasing.

Running every AI function directly on a robot therefore creates a fundamental conflict between robotic intelligence and physical resource constraints.

An alternative architecture is to regard Physical AI not as a single model located inside a robot but as a distributed intelligence system spanning the robot, communication network, edge computing infrastructure, and central AI cloud.

This transition creates an important role for 5G-Advanced and 6G networks.

2. From Monolithic VLA to Hierarchical Physical AI

Early robots receive camera images, proprioceptive measurements, or language instructions and produce an action.

Modern VLA models significantly improve this paradigm by combining visual observations and language instructions with learned action generation. Nevertheless, single-model architecture encounters difficulties when robots must simultaneously support:

- millisecond-level stabilization
- collision avoidance
- dexterous manipulation
- semantic reasoning
- long-horizon planning
- memory
- environmental prediction
- simulation
- continuous learning

These functions operate at dramatically different time scales.

For example, motor control may operate at hundreds or thousands of Hertz, while a semantic planner may only need to update several times per second. A world model used for long-horizon prediction may operate even more slowly but require orders of magnitude greater computing capacity.

This motivates a hierarchical Physical AI architecture.

3. Proposed System 0–3 Physical AI Model

Physical AI is represented using four functional layers.

3.1 System 0 – Physical perception and reflex intelligence

System 0 represents the closest intelligence layer to physical hardware.

Typical functions include:

- joint and actuator control
- force and torque processing
- tactile processing and contact detection
- proprioception
- balance control
- emergency collision avoidance
- safety interlocks

Typical inputs include:

- vision encoder
- IMUs
- tactile arrays
- force/torque sensors
- motor current
- proximity sensors
- high-frequency local perception

System 0 operates at approximately sub-millisecond to tens-of-milliseconds time scales, depending on the function.

Because communication failure must not destabilize a robot, System 0 should normally remain on-device.

3.2 System 1 – Fast perception-action intelligence

System 1 represents fast learned behavior.

Typical components may include:

- VLA action heads
- action experts
- learned motor policies
- diffusion-based action generation
- flow-matching action models
- trajectory generation
- short-horizon motion policies

System 1 transforms interpreted sensory information and task context into actions or short action sequences.

Its characteristic time scale can range approximately from 10 ms to 100 ms, although some robotic control functions will require faster local execution.

System 1 should therefore also predominantly be executed locally.

However, unlike System 0, selected System 1 functions may benefit from network-edge acceleration when communication and compute conditions satisfy the application's deadline and safety requirements.

3.3 System 2 – Reasoning and planning intelligence

System 2 provides higher-level cognitive capabilities.

Potential components include:

- Vision-Language Models
- multimodal foundation models
- semantic reasoning
- chain-of-thought-like internal planning mechanisms
- task decomposition
- long-horizon planning
- spatial reasoning
- memory retrieval
- multi-agent coordination
- semantic scene understanding

A System 2 process might receive the command:

"Move the container from the loading area to workstation B, but avoid the human working in aisle 3."

System 2 could convert this instruction into semantic subtasks:

1. identify the target container
2. determine its current location
3. evaluate environmental constraints
4. select a feasible route
5. determine an appropriate manipulation strategy
6. generate subgoals for System 1
7. monitor task completion

Unlike System 0 and most System 1 functions, System 2 does not necessarily need to execute at motor-control frequency.

Its useful response time may range approximately from tens of milliseconds to several hundred milliseconds, and in some deliberative applications longer response times may be acceptable.

At the same time, System 2 can require substantial GPU/NPU capacity.

System 2 therefore represents the primary candidate for network-edge deployment.

3.4 System 3 – World intelligence and long-term learning

System 3 represents large-scale intelligence extending beyond immediate robot behavior.

It may include:

- World Models and World Foundation Models

- digital twins
- large-scale multimodal memory
- long-horizon prediction
- physics-aware simulation
- fleet learning
- synthetic-data generation
- policy evaluation
- reinforcement learning
- model training and fine-tuning
- collective knowledge across multiple robots.

A System 3 model can answer questions such as:

"What is likely to happen over the next several minutes if the robot follows strategy A instead of strategy B?"

or:

"What manipulation policy should be learned from the experiences of 10,000 robots operating during the previous month?"

System 3 may require extremely large computational and storage resources but often has less stringent real-time requirements.

Consequently, System 3 is naturally suited to:

- regional AI data centers
- sovereign AI infrastructure
- Industrial AI clouds
- hyperscale AI clouds

4. Distributed Physical AI Architecture

The four layers exhibit substantially different requirements.

Table 1. Components of distributed physical AI architecture

Physical AI layer	Primary role	Indicative response timescale	Relative compute	Typical placement
System 0	Reflex, sensing, safety, control	<1–20 ms	Low–Medium	Robot
System 1	Fast learned action	~10–100 ms	Medium–High	Robot / local edge
System 2	Reasoning and planning	~30–500+ ms	High–Very High	Robot / MEC / edge cloud
System 3	World modeling, simulation, learning	~0.1 s–minutes+	Very High–Extreme	Edge / regional / central cloud

These numbers represent architectural design ranges rather than universal requirements. Actual limits depend strongly on robot type, velocity, safety requirements, task complexity, model architecture, and deployment environment.

A distributed architecture can therefore exploit the relationship:

Lower layer → tighter latency → closer to robot

Higher layer → larger computation → greater opportunity for remote execution

This produces a natural compute hierarchy:

Robot → Access Edge → MEC → Regional AI Cloud → Central AI Cloud

The Network-Enabled Distributed Physical AI architecture shown conceptually below.

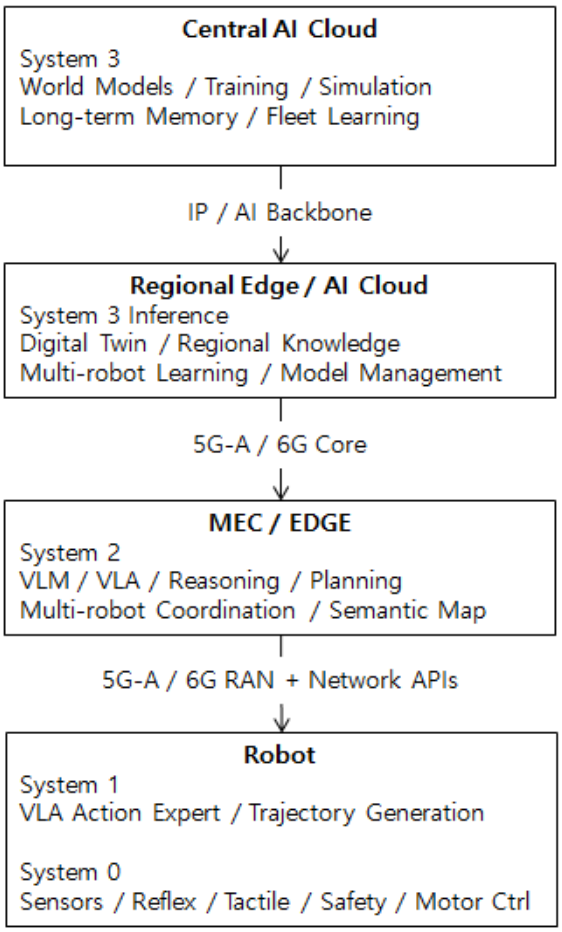


Fig. 1. Distributed physical AI architecture

The architecture differs fundamentally from conventional cloud robotics. In conventional cloud robotics, the network primarily connects a robot to a remote computing service.

In NDPA, the network participates in deciding where intelligence executes. The network therefore becomes part of the Physical AI runtime.

5. Communication Requirements

5.1 System 0 communication

System 0 should primarily use internal robotic communication.

Examples include:

- EtherCAT
- CAN/CAN-FD
- Time-Sensitive Networking
- proprietary real-time buses
- local high-speed sensor interfaces

Network communication can nevertheless expose selected System 0 information for:

- monitoring
- digital twins
- predictive maintenance
- telemetry
- anomaly detection

The network must not normally close safety-critical motor loops over a public or non-deterministic infrastructure.

5.2 System 1 communication

System 1 introduces more interesting network requirements.

Rather than transmitting individual motor commands, a distributed architecture should preferably exchange semantic actions or intermediate representations.

For example, System 2 could transmit:

```
Task:
    Pick(object_37)

Target pose:
    x, y, z, roll, pitch, yaw

Constraints:
max_force = ...
avoid_region = ...
grasp_type = ...

Confidence:
    0.93
```

System 1 can then locally convert this semantic goal into a high-frequency action trajectory.

This significantly reduces bandwidth and isolates the fast control loop from network jitters.

A key architectural principle is therefore:

The System 1 and System 2 interface should exchange intentions, subgoals, embeddings, scene representations, or action chunks rather than individual actuator commands whenever possible.

5.3 System 2 communication

System 2 creates perhaps the strongest opportunity for telecommunications infrastructure.

Large multimodal reasoning models may require substantially more compute and memory than can economically be integrated into every robot.

A distributed architecture can create a shared GPU/NPU pool at the edge.

This provides several potential benefits:

- reduced robot hardware cost
- lower robot power consumption and longer battery life
- simplified AI hardware upgrades
- centralized model lifecycle management
- shared foundation models
- improved cross-robot knowledge
- higher utilization of expensive AI accelerators

The MEC therefore becomes an AI inference fabric for Physical AI.

6. Indicative Network Requirements

A single bandwidth or latency requirement cannot represent Physical AI.

Instead, network requirements should be differentiated by AI function.

Table 2. Network performance requirements per function

Function	Indicative latency objective	Bandwidth characteristic	Reliability importance
System 0 telemetry	1–20 ms where real-time monitoring is needed	Low–Medium	Very high
System 1 edge assistance	10–100 ms	Medium–High	Very high
System 2 reasoning	30–500 ms	Medium–Very high	High
Remote perception	10–100 ms	Very high	Very high
World-model inference	100 ms–seconds	High–Extreme	Medium–High

Function	Indicative latency objective	Bandwidth characteristic	Reliability importance
Digital-twin synchronization	10 ms–seconds	Medium–Very high	High
Fleet learning	seconds–hours	Very high aggregate volume	Medium
Model distribution	seconds–minutes	Very high downlink volume	High
Teleoperation	application dependent; potentially <10–50 ms RTT for tightly coupled interaction	High uplink + downlink	Extremely high

Again, these values should not be interpreted as standardized requirements. They define candidate engineering regions for system design.

Future 6G Physical AI networks should jointly optimize this decision.

7. Role of 5G-Advanced

5G-Advanced provides an important transition platform toward distributed Physical AI.

Relevant capabilities include:

7.1 MEC

Multi-access Edge Computing places GPU/NPU infrastructure closer to robots.

This enables latency-sensitive System 2 inference and selected System 1 acceleration.

7.2 Network slicing and differentiated QoS

Physical AI traffic has heterogeneous requirements.

A robot may simultaneously generate:

- safety telemetry
- AI inference requests
- video streams
- digital-twin updates
- background training data

QoS mechanisms can prioritize latency- and reliability-sensitive robotic traffic over delay-tolerant model or training transfers.

7.3 Network exposure APIs

Network APIs can expose communication conditions to Physical AI orchestration.

Examples include:

- network quality
- estimated latency
- congestion
- location
- QoS availability
- edge-compute availability

The Physical AI Orchestrator can use this information when deciding whether an AI workload should be executed locally or remotely.

7.4 Positioning

Network-based positioning can complement:

- GNSS
- visual SLAM
- LiDAR SLAM
- UWB
- odometry.

Position information can also become an input to System 2 and System 3 environmental models.

8. Evolution toward 6G

While 5G-Advanced enables early implementations, 6G has the opportunity to make distributed Physical AI a native network capability.

We identify six particularly important areas.

8.1 AI-native networking

6G networks can use AI not only internally for network optimization but also expose network intelligence directly to Physical AI applications.

The relationship becomes bidirectional; AI for Network and Network for AI.

For robotics, the second direction may become equally important.

8.2 Communication-compute convergence

Traditional networks allocate communication resources.

Physical AI networks must allocate both Radio resources and Computing resources

A robotic inference transaction is completed only when:

1. sensor/context information reaches the compute resource
2. an accelerator becomes available
3. inference completes
4. the result returns before the application deadline

Optimizing radio latency alone is insufficient.

The network must optimize the complete AI transaction latency.

9. Integrated Sensing and Communication

6G Integrated Sensing and Communication (ISAC) can provide another important Physical AI capability.

Network infrastructure can potentially sense:

- object presence
- movement
- velocity
- trajectories
- environmental changes

This creates a new perception architecture.

The network becomes a supplementary sensor for the robot, which is particularly valuable when objects are:

- outside the robot's field of view
- occluded
- behind corners
- beyond local sensing range
- observed by other network-connected agents

10. Semantic and Goal-Oriented Communication

Future Physical AI networks should not necessarily transmit every bit produced by every sensor.

Communication should increasingly become task aware.

Suppose System 2 only needs to know:

"Is a person entering the robot's planned path?"

Transmitting a continuous 4K video stream may be unnecessary.

A local model could instead transmit:

Human:
detected = true
position = (...)
velocity = (...)
trajectory = (...)
confidence = 0.97

This potentially reduces network load by orders of magnitude.

6G research should therefore investigate semantic communication for Physical AI, where the communication objective becomes preservation of task-relevant information rather than reconstruction of every transmitted bit.

11. World Models and Digital Twins

System 3 provides another major opportunity for telecom and cloud infrastructure.

A World Model attempts to represent the dynamics of an environment and predict future states while Digital twins provide structured virtual representations of physical systems. These technologies can converge.

System 2 can use these predictions to select better plans before System 1 executes them.

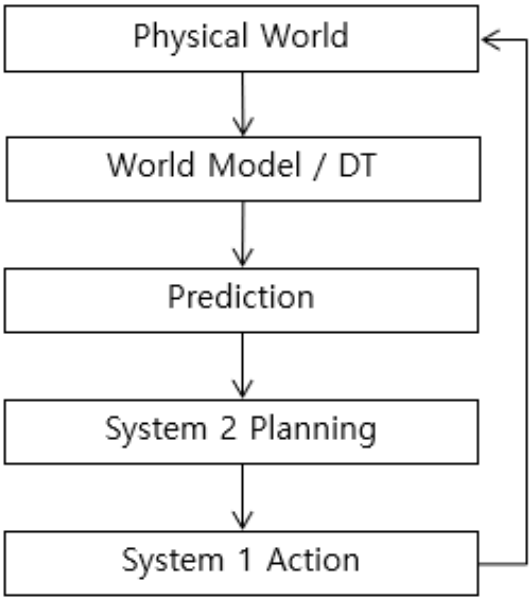


Fig. 2. Information flow diagram for the world models in loop

This is fundamentally different from a digital twin used only for offline visualization.

The digital twin becomes an active component of robotic intelligence.

12. Security and Trust Architecture

Distributing robotic intelligence across networks creates new attack surfaces.

Threats include:

- malicious AI instructions
- model substitution
- sensor-data manipulation
- adversarial observations
- robot impersonation
- compromised edge models
- unauthorized control
- semantic-data poisoning
- denial-of-service attacks against AI inference

Security therefore must protect not only packets but also the AI transaction.

Future architectures should support:

- robot identity
- workload identity
- model identity
- authenticated AI endpoints
- integrity verification
- trusted execution environments
- confidential computing
- policy-based authorization
- continuous anomaly detection

A System 2 command sent to a physical robot can cause real-world movement. Its trust requirements are consequently higher than those of conventional cloud inference.

The orchestration plane is the critical element connecting telecommunications infrastructure with Physical AI.

13. Key 6G Research Requirements

Based on this architecture, we propose that 6G Physical AI research address at least the following requirements.

Table 3. 6G network requirements for physical AI motivation

Requirement	Physical AI motivation
Deterministic low latency	Remote/edge AI must meet robotic deadlines
High uplink capacity	Robots are intensive sensor-data producers
High reliability	Missing AI responses can interrupt physical tasks
Edge GPU/NPU integration	System 2 requires large-scale inference
Compute-aware networking	Network path must consider accelerator availability
AI transaction SLA	Packet latency alone is insufficient

Requirement	Physical AI motivation
Network APIs	AI orchestrator requires network state
ISAC	Network becomes additional perception infrastructure
Precise positioning	Provides spatial context for Physical AI
Semantic communication	Reduce unnecessary raw sensor transmission
Multi-agent communication	Enables collective intelligence
Distributed AI security	AI instructions can create physical consequences
Energy-aware orchestration	Optimize robot battery and network/compute energy
Mobility-aware AI continuity	AI sessions must follow moving robots
Model/context migration	Reasoning state must move between edge locations

14. Conclusion

Physical AI is evolving toward increasingly sophisticated hierarchical intelligence. Reflexes, learned actions, reasoning, and world modeling operate at different temporal scales and require dramatically different computational resources.

Attempting to execute all these capabilities within every robot creates constraints in cost, power, thermal design, computing capacity, and model scalability.

This white paper therefore proposes a Network-Enabled Distributed Physical AI Architecture in which Physical AI is distributed across robots, 5G/6G networks, edge computing infrastructure, and central AI clouds.

The architecture maps System 0 predominantly to the robot, System 1 primarily to local computing, System 2 to a flexible robot–MEC continuum, and System 3 predominantly to edge and cloud AI infrastructure.

The central proposition is that future networks should not merely transport robotic data.

They should become an integral component of robotic intelligence.

This requires telecommunications research to move beyond conventional KPIs such as throughput and packet latency toward concepts including:

- AI transaction latency;
- compute-aware networking;
- dynamic AI workload placement;
- semantic robotic communication;
- network-assisted perception;
- collective Physical AI;
- AI workload continuity;
- edge foundation-model inference; and
- joint communication-computation-energy optimization.

5G-Advanced provides many of the initial building blocks through MEC, advanced QoS, network APIs, positioning, and private networking. 6G provides an opportunity to integrate communication, computing, sensing, and AI much more deeply.

The resulting architecture represents a transition from networks connecting robots to networks becoming part of robots' intelligence.

Ultimately, this suggests a broader vision for 6G rather than simply a faster communication network.

References

1. NVIDIA: Cosmos World Foundation Model Platform for Physical AI (2025)
2. NVIDIA: NVIDA Isaac GR00T: Generalist Robot 00 Technology (2025)
3. Google: Physical AI: The Next Frontier in AI and Robotics to Build Truly Autonomous Machines (2026)
4. Google: Shaping the future of advanced robotics (2024)
5. Indian Institute of Technology Bombay: AI/ML in 3GPP 5G Advanced – Services and Architecture (2025)
6. Ericsson: 5G Advanced Evolution towards 6G (2023)
7. 3GPP: WG CT3 Rel-19 AI/ML features (2026)
8. ITU-T: IMT-2030: Technical requirements for the 6G future (2025)
9. ACM SIGCOMM Workshop: Pathways towards Network-as-a-Service: the CAMARA project (2026)

Benchmarking Oversampling Techniques and Implementing a Binary Classifier for Identifying RuBisCO Protein Sequences

Ji Sun Kwon[1], Sonja Diedrich[1], and Jochen Blom[1]

[1]Bioinformatics and Systems Biology, Justus-Liebig-Universität Gießen, Ludwigsplatz 13-15, 35390 Gießen, Germany
klara.kwon@web.de

Abstract. Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is a key enzyme of photosynthesis and thus plays a key role in carbon fixation. Nonetheless, publicly available RuBisCO sequences are scarce compared with non-RuBisCO proteins and closely related RuBisCO-like proteins (RLPs). A classification of novel RuBisCO proteins using modern Machine Learning approaches would thus be highly desirable, but the imbalance of the data makes this a difficult task. To address this problem, this study systematically benchmarked three oversampling techniques in this context: Random Oversampling (RO), Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN). Three datasets containing RuBisCO, non-RuBisCO and optionally RLP sequences were built in five sizes (8,910 to 29,478 entries) and used to train four machine-learning and deep-learning models: Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Multi-Layer Perceptron (MLP), evaluated by F1-score, Receiver Operating Characteristic Curve - Area Under the Curve (ROC-AUC) and Matthews Correlation Coefficient (MCC) alongside computational cost. Baseline models without oversampling already exceeded 0.98 accuracy, leaving little room for improvement. Oversampling produced only marginal gains (below 0.003) and occasionally reduced performance, notably ADASYN on larger datasets. RF was most stable with RO, whereas SVM and MLP benefited most from SMOTE. Including RLPs slightly lowered accuracy, reflecting their similarity to RuBisCO. Computational cost grew nearly linearly with data size, steepest for SVM and XGBoost. The results highlight how biological bias, model-specific behavior and computational demand shape the value of oversampling in bioinformatics.

Keywords: Data Imbalance, RuBisCO, Binary Classification, Oversampling (Random Oversampling, SMOTE, ADASYN), Machine Learning (Random Forest, SVM, XGBoost), Deep Learning (MLP).

1. Introduction

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is the most abundant enzyme on Earth and is central to photosynthesis and carbon fixation, assimilating roughly 250 billion tonnes of CO2 into the biosphere each year. This makes it highly relevant to climate change and crop yield. Reliably distinguishing RuBisCO from non-RuBisCO proteins and from RuBisCO-like proteins (RLPs) is, however, difficult because of their structural similarity [5], providing a biologically meaningful yet technically challenging classification task.

Like many bioinformatics datasets, RuBisCO versus non-RuBisCO data are strongly imbalanced: non-RuBisCO proteins are abundant while curated RuBisCO sequences are comparatively rare. Such imbalance biases classifiers toward the majority class, producing misleadingly high overall accuracy while missing true RuBisCO proteins. Oversampling is a widely used remedy for class imbalance in health and medical diagnostics, but comprehensive evaluations are recent [1] and its effectiveness has not been examined for RuBisCO classification.

Sequence-alignment tools such as BLAST remain useful but rely mainly on local similarity [13], and high sequence identity does not always imply functional or structural similarity. This limitation is particularly relevant to RuBisCO and RLPs, which share high similarity yet differ in catalytic activity. Machine-Learning (ML) and Deep-Learning (DL) models, combined with oversampling, can potentially capture sequence patterns beyond direct similarity.

This study therefore benchmarks three representative oversampling techniques: Random Oversampling (RO), Synthetic Minority Oversampling Technique (SMOTE) [2] and Adaptive Synthetic Sampling (ADASYN) [3], using RuBisCO versus non-RuBisCO classification as a case study. The aims are to determine whether oversampling improves minority-class (RuBisCO) detection, how technique choice interacts with model type and dataset size, and to derive broader guidance for handling class imbalance in biological sequence data.

2. Experiments and Discussion

2.1 Datasets and Experimental Workflow

Three datasets were assembled. The positive set comprised 3,768 large-chain bacterial RuBisCO sequences retrieved from UniProt [7] (136 reviewed Swiss-Prot, 3,632 unreviewed TrEMBL); large chains were chosen because they represent the catalytically active form and their selection automatically excludes short-chain RLPs. The negative set of non-RuBisCO proteins was sampled from UniRef90 clusters using cluster-stratified sampling to maximise diversity, in five sizes (5,000–25,000). A separate RLP set of 852 sequences was drawn from BRENDA/UniProt. Two experiments were defined: Experiment 1 (positive + negative, five sizes 8,768–28,768) and Experiment 2 (positive + negative + RLP, five sizes 8,910–29,478). Larger datasets were progressively more imbalanced, from 1:1.3 up to 1:6.6.

A cross-validation (CV) pipeline was implemented so that all preprocessing occurred inside the CV loop to prevent data leakage. After stratified K-fold splitting, cleaning (median/most-frequent imputation, duplicate and outlier removal beyond the 99.5% length quantile) and encoding were applied. High-cardinality metadata were hash- or multi-label-encoded, while each protein sequence was encoded with the BLOSUM62 substitution matrix [6], mapping each residue to its 24-dimensional row vector and flattening the result into a fixed-length feature vector. This biologically informed encoding preserves substitution similarity, unlike one-hot encoding. Sequence length was standardised with StandardScaler.

Oversampling (RO, SMOTE, ADASYN) was then applied to the training folds only, with hyperparameters (sampling ratio, number of neighbors) tuned via GridSearchCV using imbalanced-learn [4]. Four classifiers were trained: Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost) and a Multi-Layer Perceptron (MLP), and evaluated with F1-score, Receiver Operating Characteristic Curve - Area Under the Curve (ROC-AUC) and the Matthews Correlation Coefficient (MCC), metrics chosen for their complementary behaviour under imbalance [11]. Jobs were executed on a SLURM cluster, and resource usage was profiled to quantify computational cost.

2.2 Classification Performance

Across both experiments, baseline models without any oversampling already achieved near-maximal performance: ROC-AUC typically exceeded 0.99 and F1/MCC reached 0.98–0.99. This is best read as a success of the study, not a limitation: it shows that the rigorous, leakage-free preprocessing and biologically informed BLOSUM62 encoding established in Section 2.1 alone made RuBisCO and non-RuBisCO sequences almost perfectly separable, without any class-balancing. Rather than making oversampling redundant, this near-ceiling baseline gave a stringent, negative-control-like reference against which the true, isolated effect of oversampling could be measured. Against it, oversampling produced only third-decimal changes: RO and SMOTE gave small, consistent gains, whereas ADASYN occasionally fell below baseline, most clearly on larger datasets. RF was most stable with RO, consistent with its robustness to duplicated samples, while SVM and MLP benefited most from SMOTE, whose interpolation smooths decision boundaries. ADASYN improved minority recall in the smallest datasets, where imbalance was most acute, but did not scale well. Including RLPs (Experiment 2) slightly lowered scores, especially for the smallest datasets, confirming that their close similarity to RuBisCO makes classification harder; this gap narrowed as sample size increased.

Oversampling hyperparameters had only a minor influence on this highly separable dataset. RO performed best at a modest sampling ratio (0.5), SMOTE at full balancing (1.0), and ADASYN peaked near 0.9 before declining, suggesting excessive adaptive sampling introduces noise. Both SMOTE and ADASYN were best with three neighbors, and SMOTE consistently outperformed ADASYN. ROC-AUC remained above 0.998 throughout, offering little discriminating power between settings.

The most informative result came from per-class recall (Table 1). Non-RuBisCO recall was consistently near 1.0, while RuBisCO recall was about 0.05 lower and dropped further on the largest dataset. This shows that the near-perfect aggregate metrics were driven by the majority class [11], and that oversampling becomes more relevant as imbalance grows: ADASYN gave the best minority recall on the smallest dataset, whereas RO was more effective at scale.

Table 1. Per-class recall for the RF model on the smallest (n = 8,910) and largest (n = 29,478) datasets of Experiment 2 (with RLP), for the held-out test set. Table captions are placed above the table.

Oversampling	RuBisCO (n=8,910)	non-RuB. (n=8,910)	RuBisCO (n=29,478)	non-RuB. (n=29,478)
None	0.9648	1.0000	0.9391	1.0000
RO	0.9662	1.0000	0.9567	0.9998
SMOTE	0.9662	1.0000	0.9513	1.0000
ADASYN	0.9675	1.0000	0.9445	0.9998

A separability test trained a logistic-regression classifier to distinguish real from synthetic minority samples. An AUC near 0.5 means synthetic data are indistinguishable from real data. SMOTE achieved the lowest AUC (0.704), followed by RO (0.723), while ADASYN reached 0.976. This mechanistically explains ADASYN's instability: by focusing on difficult boundary regions, it generated scattered, less realistic points that expanded feature space and degraded generalisation on larger datasets, as also seen in PCA projections.

Computational cost grew almost linearly with dataset size for all models but differed by architecture. SVM and XGBoost were most demanding, XGBoost exceeded 120 GB at the largest size, whereas RF and MLP remained lower and similar. These patterns match the expected scaling of each model, with SVM's cost between $O(n^2)$ and $O(n^3)$ explaining its steep growth. Oversampling amplified cost non-uniformly: negligible at the smallest size but substantial for XGBoost and MLP at the largest, since synthetic samples inflate the training set and, for MLP, are reprocessed every epoch. RO scaled most predictably as it merely duplicates points.

Biologically, the limited benefit of oversampling reflects the dataset itself. The positive class was highly homogeneous (dominated by RuBisCO large-subunit entries, typically 450–500 aa) while the negative class was heterogeneous and shorter (400–440 aa), giving models a strong length signal for near-perfect baseline separation even without resampling. However, most entries were unreviewed and inferred from homology, so classification was partly driven by annotation and collection biases rather than biology alone [10], which limits generalisation. RuBisCO classification is thus not an especially challenging test for oversampling, consistent with evidence that certain protein families are readily distinguished from sequence data alone [8],[9], and its value should be clearer for less distinctive protein families or noisier labels.

Model-specific behaviour was consistent with theory. RF, relying on bootstrapped feature splits and majority voting, is robust to duplicates and gains little from synthetic interpolation. SVM and MLP, being boundary-sensitive, benefited from SMOTE's smooth interpolation, while ADASYN's outliers occasionally harmed MLP generalisation. XGBoost gained from SMOTE filling minority gaps but was destabilised by ADASYN. These findings agree with prior work reporting that strong baselines can leave oversampling gains negligible and that no single resampling method dominates [12], which also helps explain why undersampling has historically been preferred in bioinformatics.

3. Conclusions

This study provides the first systematic benchmark of RO, SMOTE and ADASYN for RuBisCO versus non-RuBisCO classification across four ML/DL models, five dataset sizes and two dataset compositions. Because BLOSUM62-encoded sequences and a strong length signal made the classes highly separable, baseline models already reached above 0.98 accuracy and oversampling added little (below 0.003), sometimes reducing performance. The main contribution is therefore a nuanced one: technique choice still matters, RO suits RF, SMOTE suits SVM and MLP, and ADASYN helps only under strong imbalance in small datasets, and per-class recall reveals that aggregate metrics hide a persistent minority-class gap that widens with imbalance.

Beyond performance, the work shows that the usefulness of oversampling in bioinformatics depends jointly on biological data traits, model-specific behaviour and computational feasibility, since synthetic generation can substantially inflate cost for SVM, XGBoost and MLP at scale. Future work should test more imbalanced or less separable datasets, richer RLP representation and hybrid or cluster-based oversampling variants, together with metrics that better quantify how faithfully synthetic samples reproduce real biological sequences.

References

1. V. Werner de Vargas, J. A. Schneider Aranda, R. dos Santos Costa, P. R. da Silva Pereira, and J. L. Victória Barbosa, "Imbalanced data preprocessing techniques for machine learning: a systematic mapping study," Knowledge and Information Systems, vol. 65, no. 1, pp. 31–57, Jan. 2023, doi: 10.1007/s10115-022-01772-8
2. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.
3. H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in Proc. IEEE Int. Joint Conf. Neural Networks (IJCNN), Hong Kong, China, 2008, pp. 1322-1328.
4. G. Lemaitre, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," Sep. 2016, [Online]. Available: http://arxiv.org/abs/1609.06570.
5. F. R. Tabita, T. E. Hanson, H. Li, S. Satagopan, J. Singh, and S. Chan, "Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships," J. Exp. Bot., vol. 58, no. 13, pp. 3093–3100, 2007, doi: 10.1093/jxb/erm204.
6. S. R. Eddy, "Where did the BLOSUM62 alignment score matrix come from?," Nature Biotechnology, vol. 22, no. 8, pp. 1035–1036, Aug. 2004, doi: 10.1038/nbt0804-1035.
7. A. Bateman et al., "UniProt: the Universal Protein Knowledgebase in 2023," Nucleic Acids Research, vol. 51, no. D1, pp. D523–D531, Jan. 2023, doi: 10.1093/nar/gkac1052.
8. C. Jin and R. I. Cukier, "Machine learning can be used to distinguish protein families and generate new proteins belonging to those families," Journal of Chemical Physics, vol. 151, no. 17, Nov. 2019, doi: 10.1063/1.5126225.
9. W. A. Iqbal, A. Lisitsa, and M. v. Kapralov, "Predicting plant Rubisco kinetics from RbcL sequence data using machine learning," Journal of Experimental Botany, vol. 74, no. 2, pp. 638–650, Jan. 2023, doi: 10.1093/jxb/erac368.
10. A. M. Schnoes, D. C. Ream, A. W. Thorman, P. C. Babbitt, and I. Friedberg, "Biases in the Experimental Annotations of Protein Function and Their Effect on Our Understanding of Protein Function Space," PLoS Computational Biology, vol. 9, no. 5, 2013, doi: 10.1371/journal.pcbi.1003063.
11. G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," PLoS ONE, vol. 18, no. 10, Art. no. e0291908, Oct. 2023. doi: 10.1371/journal.pone.0291908. [Online]. Available: https://doi.org/10.1371/journal.pone.0291908.
12. D. J. Dittman, T. M. Khoshgoftaar, R. Wald, and A. Napolitano, "Comparison of Data Sampling Approaches for Imbalanced Bioinformatics Data." [Online]. Available: www.aaai.org.
13. A. Al-Fatlawi, M. Menzel, and M. Schroeder, "Is Protein BLAST a thing of the past?," Nature Communications, vol. 14, no. 1. Nature Research, Dec. 01, 2023. doi: 10.1038/s41467-023-44082-5

Topological Features for Environmental Sound Classification: A Proof of Concept toward Industrial Acoustic Anomaly Detection[1]

Jiwon Seo[1], Sascha Grollmisch[1], Jaehyun Lee[2], Sukjin Hong[2], and Jooyoung Hahn[3]

[1] Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany
[2] LOAS Inc., Seoul, Republic of Korea
[3] Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic
jiwon.seo@idmt.fraunhofer.de

Abstract. This paper summarizes our previous study [1], in which we investigated whether topological features derived from persistent homology of Takens embeddings can complement conventional acoustic features for environmental sound classification on the ESC-50 dataset. A compact statistical summary of persistence diagrams is extracted from short overlapping segments, aggregated at the recording level, and combined with conventional acoustic features. Experiments show that fusing topological and acoustic descriptors consistently improves classification performance across multiple classifiers. With a Random Forest classifier, the proposed fusion achieves 48.5% accuracy, an absolute improvement of 7.6 percentage points over the acoustic baseline (MFCC + ZCR). With a multilayer perceptron, the improvement reaches 12.1 percentage points. A dimensionality-matched control experiment confirms that these gains arise from informational content, not the extended dimensions of the fusion approach.

Keywords: topological data analysis, persistent homology, environmental sound classification, feature complementarity, acoustic anomaly detection.

1. Introduction

Acoustic monitoring is increasingly used to detect faults and anomalies in industrial machinery. This work is part of a Korea–Germany research collaboration between Fraunhofer IDMT and LOAS Inc. on anomaly detection in noisy industrial acoustic signals. As a controlled proof of concept, we study whether topological features complement conventional acoustic features on the ESC-50 environmental-sound benchmark [2]. Conventional approaches rely on acoustic descriptors derived from time-frequency representations, such as mel-frequency cepstral coefficients (MFCCs) [3]. Within the conventional feature-based framework, the question of whether descriptors derived from the time-domain signal structure can complement established acoustic features such as MFCCs remain open and is the focus of this work.

Topological Data Analysis (TDA) offers methods to analyze complex data by focusing on the topological structure of the data. One of its central tools, persistent homology, tracks how topological structures, such as connected components and loops, appear and disappear across a range of scales [4]. Applied to time-series data via Takens embeddings [5], persistent homology can reveal structural patterns that conventional signal representations do not capture. Most relevant to our work, Fireaizen et al. [6] combined persistent homology of time-domain embeddings with acoustic features such as MFCCs and reported that this fusion substantially improved alarm detection over acoustic features alone. That study, however, addressed only a single binary task, the detection of alarm sounds.

This paper summarizes our Forum Acusticum 2026 study [1]. We investigated whether a simple topological descriptor, a compact statistical summary of persistence diagrams obtained from Takens embeddings of time-domain audio, could complement conventional acoustic features, namely MFCCs, for environmental sound classification on the ESC-50 dataset. Our goal was not to maximize absolute accuracy or to benchmark alternative vectorization schemes, but to establish whether persistent homology provides complementary information to standard acoustic features, which we tested with a dimensionality-matched control and a feature-importance analysis.

[1]This article is a condensed summary of work presented at Forum Acusticum 2026 [1].

2. Method

Figure 1 outlines the processing pipeline. Experiments were conducted on the ESC-50 dataset [2], which contains 2,000 environmental audio recordings organized into 50 balanced sound event classes (e.g., dog, rain, siren, and sneezing), following the official 5-fold cross-validation protocol. To obtain stable topological descriptors, feature extraction was performed on short overlapping segments: each recording was divided into 1 s segments with a 0.5 s hop size, yielding nine segments per recording.

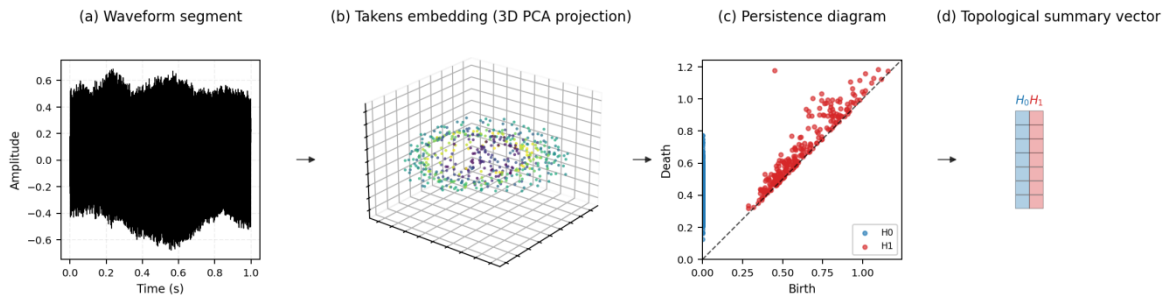


Fig. 1. Topology-based feature extraction: a waveform segment is mapped to a point cloud by a Takens embedding, summarized by the persistent homology of the point cloud in dimensions H_0 and H_1, and reduced to a fixed-length statistical descriptor. Reproduced from [1].

Topological features. A Takens embedding [5] maps each segment to a point cloud by sliding a short window of consecutive samples along the waveform and treating each window as one point, so that temporal patterns in the signal appear as geometric structure (embedding dimension d = 32, delay τ = 1 sample, stride 16 samples). A Vietoris–Rips filtration [7] then connects points closer than a growing distance threshold; as the threshold increases, connected components progressively merge, while loops appear and close. Each connected component (H_0) and loop (H_1) is recorded in a persistence diagram by its birth and death threshold, and the difference between them, the persistence, indicates how strongly the structure persists. Since the diagram has a variable number of points, the persistence values are summarized by seven statistics (count, sum, mean, standard deviation, and the three largest values) for each of H_0 and H_1, giving a fixed 14-dimensional descriptor. Mean and standard-deviation pooling across segments produced a 28-dimensional recording-level topological vector, PH(H_0,H_1).

Acoustic features. For each segment, 12 mel-frequency cepstral coefficients ($MFCC_{12}$) and the zero-crossing rate (ZCR) were computed following the baseline of [2], yielding a 13-dimensional segment-level descriptor. Mean and standard deviation pooling produced a 26-dimensional recording-level feature vector, denoted $MFCC_{12}$ + ZCR. To match the dimensionality of the proposed fusion, we additionally computed an extended acoustic set, $MFCC_{26}$ + ZCR, using 26 coefficients (54 dimensions after pooling).

Fusion and classification. The acoustic and topological vectors were concatenated and standardized per feature block using z-score normalization, with statistics estimated from the training folds. We evaluated a Random Forest (RF; 1000 trees) and a multilayer perceptron (MLP; two hidden layers of 256 and 128 units), reporting mean accuracy over the official 5-fold cross-validation.

3. Results and Discussion

Table 1. Classification accuracy (%) on ESC-50 (mean over the official 5-fold cross-validation) for a Random Forest (RF) and a multilayer perceptron (MLP). Dim is the recording-level feature dimensionality. The best result in each column is in bold. Reproduced from [1].

Feature set	Dim	RF	MLP
$MFCC_{12}$ + ZCR (acoustic baseline)	26	40.9	43.0
$MFCC_{26}$ + ZCR (dim.-matched control)	54	44.4	44.5
PH(H_0,H_1) (topological only)	28	33.3	36.3
$MFCC_{12}$ + ZCR + PH(H_0,H_1) (fusion)	**54**	**48.5**	**55.1**

Topological features alone yielded lower but non-trivial performance, with PH(H_0,H_1) reaching 33.3% (RF) and 36.3% (MLP). Fusing topological and acoustic features consistently improved performance across both classifiers: the best results were obtained with $MFCC_{12}$ + ZCR + PH(H_0,H_1), reaching 48.5% with RF and 55.1% with MLP, corresponding to absolute improvements of 7.6 and 12.1 percentage points over the respective baselines.

To verify that the performance gains were not attributable to the increase in feature dimensionality alone, we compared the proposed fusion $MFCC_{12}$ + ZCR + PH(H_0,H_1) against the control $MFCC_{26}$ + ZCR, both 54-dimensional. While $MFCC_{26}$ + ZCR improved over $MFCC_{12}$ + ZCR by 3.5 (RF) and 1.5 (MLP) percentage points, the proposed fusion yielded substantially larger gains of 7.6 (RF) and 12.1 (MLP) percentage points. The gap between the two 54-dimensional configurations indicates that topological descriptors contribute information beyond what additional MFCCs provide.

Table 2. Feature-group importance for the fused model on ESC-50. MDI is the mean decrease in impurity, which sums to one across the groups; Perm. is the drop in accuracy when a group is jointly permuted. The topological row is in bold. Reproduced from [1].

Feature group	RF MDI	RF Perm.	MLP Perm.
$MFCC_{12}$ + ZCR (acoustic)	0.51	0.29	0.45
PH(H_0,H_1) (topological)	**0.49**	**0.36**	**0.37**

A feature-importance analysis further supported this (Table 2). The topological descriptors held 49% of the Random Forest mean decrease in impurity (MDI), and jointly permuting them lowered accuracy by 0.36 (RF) and 0.37 (MLP), about twice the drop from the higher-order MFCCs added by the control (0.18 and 0.19). Both classifiers therefore relied substantially on the topological descriptors.

4. Conclusion and Outlook

In the study summary here [1], we investigated whether topological features from persistent homology of Takens embeddings complement conventional acoustic features, namely MFCCs, for environmental sound classification on the ESC-50 dataset. Experiments demonstrated consistent classification improvements across both a Random Forest and a multilayer perceptron classifier, with absolute accuracy gains of 7.6 and 12.1 percentage points over the MFCC baseline. A dimensionality-matched control experiment confirmed that these gains arise from the informational content of the topological descriptors rather than from the extended feature dimensionality. Full experimental details are reported in [1].

Within the broader Korea–Germany project, this controlled result supports the use of topological features for the target application, acoustic anomaly detection in noisy industrial machinery. Industrial machine sounds are typically periodic or quasi-periodic and are punctuated by impulsive events, the kinds of structure that persistent homology summarizes through H_0 and H_1, and faults often alter this structure; a topological descriptor that complements spectral features on diverse environmental sounds is therefore a promising building block for detecting such deviations in machine audio. The collaboration combines Fraunhofer IDMT's expertise in industrial sound analysis with LOAS Inc.'s acoustic inspection across Korean manufacturing sites, and ongoing work integrates the two feature types and evaluates them on manufacturing-site recordings.

Acknowledgements. This work was conducted as part of a joint research project between Fraunhofer IDMT and LOAS Inc., funded by the Korea Technology and Information Promotion Agency for SMEs (TIPA) under Grant No. RS-2025-25440174.

Declaration of AI use. During the preparation of this work, the authors used Claude (Anthropic) to draft, condense, and edit text based on their prior publication [1]. The authors reviewed the content and take full responsibility for it.

References

1. Seo, J., Grollmisch, S., Lee, J., Hong, S., Hahn, J.: Topological Features for Environmental Sound Classification. In: Forum Acusticum 2026, Graz, Austria (2026, to appear).
2. Piczak, K.J.: ESC: Dataset for Environmental Sound Classification. In: Proc. 23rd ACM International Conference on Multimedia, pp. 1015–1018 (2015).
3. Logan, B.: Mel Frequency Cepstral Coefficients for Music Modeling. In: International Symposium on Music Information Retrieval (2000).
4. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological Persistence and Simplification. In: Proc. 41st Annual Symposium on Foundations of Computer Science, pp. 454–463 (2000).
5. Takens, F.: Detecting Strange Attractors in Turbulence. In: Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics, vol. 898, pp. 366–381. Springer, Heidelberg (1981).
6. Fireaizen, T., Ron, S., Bobrowski, O.: Alarm Sound Detection Using Topological Signal Processing. In: ICASSP 2022, pp. 211–215 (2022).
7. Bauer, U.: Ripser: efficient computation of Vietoris–Rips persistence barcodes. Journal of Applied and Computational Topology 5(3), 391–423 (2021).

Enhanced Operational Characteristics of AlGaN/GaN HEMTs Using HfO_2-Based Passivation: A Simulation Study

Junho Lee[1] and Hyun-Seok Kim[1]

[1] Division of Electronics and Electrical Engineering, Dongguk University-Seoul, Seoul 04620, Republic of Korea
steve1211@dgu.ac.kr; hyunseokk@dongguk.edu

Abstract. This study investigates the operational characteristics of AlGaN/GaN high-electron-mobility transistors (HEMTs) using HfO_2-based passivation. A technology computer-aided design (TCAD) model was first calibrated against measured DC and RF data from a fabricated Si_3N_4-passivated HEMT. Conventional Si_3N_4, full HfO_2, and HfO_2/Si_3N_4 hybrid passivation structures were then compared. Full HfO_2 passivation redistributed the channel electric field and improved breakdown voltage by approximately 19%, but its high dielectric constant increased parasitic capacitances and significantly degraded frequency characteristics. To mitigate this trade-off, the second Si_3N_4 thickness of the hybrid structure was varied from 150 to 450 nm. A hybrid structure with a 50-nm HfO_2 first layer and a 350-nm Si_3N_4 second layer achieved a breakdown voltage of 267.57 V, cut-off frequency of 22.64 GHz, and Johnson's figure-of-merit of 6.06 THz·V. The results indicate that an appropriately designed hybrid passivation structure can improve high-voltage performance while preserving RF capability of HEMT.

Keywords: AlGaN/GaN HEMT; HfO_2; high-k dielectric; passivation; TCAD; breakdown voltage; RF characteristics

1. Introduction

AlGaN/GaN HEMTs have become important candidates for high-power and high-frequency electronic systems because GaN combines a wide bandgap of approximately 3.4 eV with a high critical electric field, high electron saturation velocity, and good thermal stability [1–4]. In an AlGaN/GaN heterostructure, spontaneous and piezoelectric polarization produces a high-density two-dimensional electron gas (2-DEG) at the heterointerface. This channel enables high current density and low channel resistance without conventional intentional channel doping [5].

Despite these advantages, simultaneous optimization of breakdown voltage and RF performance remains difficult. Electric-field crowding near the gate edge can limit off-state breakdown voltage, while common field-management techniques such as source- or gate-connected field plates increase the overlap between electrodes and consequently increase parasitic capacitance. Higher C_{gs} and C_{gd} reduce the current-gain cut-off frequency (f_T) and maximum oscillation frequency (f_{max}). Thus, improvements in high-voltage characteristics are frequently obtained at the expense of frequency performance.

HfO_2 is attractive as a surface passivation material because of its high dielectric constant (~25) and wide bandgap (~5.7 eV). The high permittivity can modify the electric-field distribution near the gate and field-plate edges, while the wide bandgap can help suppress leakage. However, the same high permittivity increases parasitic capacitance. The objective of this study is therefore not simply to replace Si_3N_4 with HfO_2, but to identify a passivation configuration that provides a useful balance between DC/high-voltage and RF characteristics.

2. Device Structure and Simulation Method

The reference device was a fabricated planar-gate AlGaN/GaN HEMT on a 4-inch 4H-SiC substrate. The epitaxial structure consisted of a 0.2-µm nucleation layer, a 2-µm Fe-doped GaN buffer layer, and a 25-nm $Al_{0.25}Ga_{0.75}N$ barrier layer. Ohmic contacts were formed using Ti/Al/Ni/Au metallization followed by rapid thermal annealing. The device employed a 0.15-µm gate foot, 0.5-µm gate-to-source spacing, and 1.5-µm gate-to-drain spacing. In the reference passivation structure, a 50-nm first Si_3N_4 layer and a 250-nm second Si_3N_4 layer were used together with a source-connected field plate. The fabricated device and the corresponding TCAD structure are shown in **Fig. 1**.

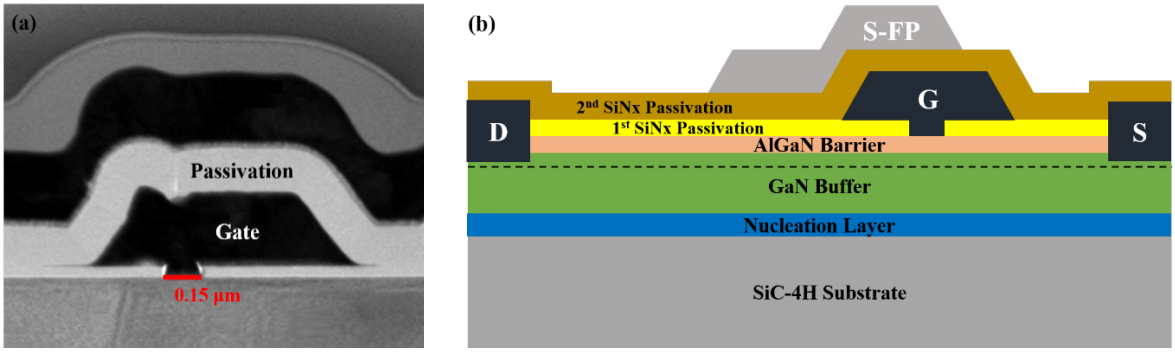


Fig. 1. Cross-sectional structure of the AlGaN/GaN HEMT: (a) SEM image of the fabricated device and (b) schematic structure used for TCAD simulation.

To establish simulation reliability before evaluating new passivation concepts, the TCAD model was calibrated to the fabricated reference device first. The simulation model included field-dependent electron mobility, Shockley–Read–Hall recombination, polarization effects, impact ionization, and self-heating. The FMCT mobility model was used in the low-field regime and the GANSAT model in the high-field regime. A Gaussian Fe acceptor-trap profile was introduced in the GaN buffer to represent semi-insulating behavior and reduce substrate leakage. The peak acceptor concentration was set in the buffer and decreased toward the AlGaN/GaN interface.

The role of passivation in RF behavior can be understood from the basic capacitance relation $C = \varepsilon_0\varepsilon_r A/d$. A material with larger relative permittivity increases capacitance when geometry is unchanged. In a HEMT, the current-gain cut-off frequency is approximately $f_T = g_m/[2\pi(C_{gs} + C_{gd})]$, showing that increased gate capacitance directly reduces f_T. The f_{max} is also sensitive to f_T, C_{gd}, and gate resistance. Therefore, the dielectric constant and physical thickness of the passivation stack are both critical design parameters.

The simulated transfer characteristics were compared with measured I_{DS}–V_{GS} data at V_{DS} = 10 V. The measured drain current at V_{GS} = 0 V was 898.71 mA/mm, while the simulation yielded 914.90 mA/mm. The measured and simulated maximum transconductance values were 344.17 and 349.60 mS/mm, respectively, with a threshold voltage of approximately −3.8 V. The maximum deviation among these principal transfer parameters was below 1.8%, indicating good DC calibration. The measured and simulated transfer characteristics are compared **in Fig. 2(a)**.

RF calibration was performed at V_{DS} = 20 V and V_{GS} = −2.6 V. The measured f_T, extracted from the −20 dB/decade current-gain slope, was 25.19 GHz, compared with a simulated value of 24.64 GHz. The error was approximately 2.2%. Because both DC and RF characteristics were reproduced with small errors, the calibrated model was considered sufficiently reliable for comparative evaluation of alternative passivation structures. The corresponding RF comparison is shown in **Fig. 2(b)**.

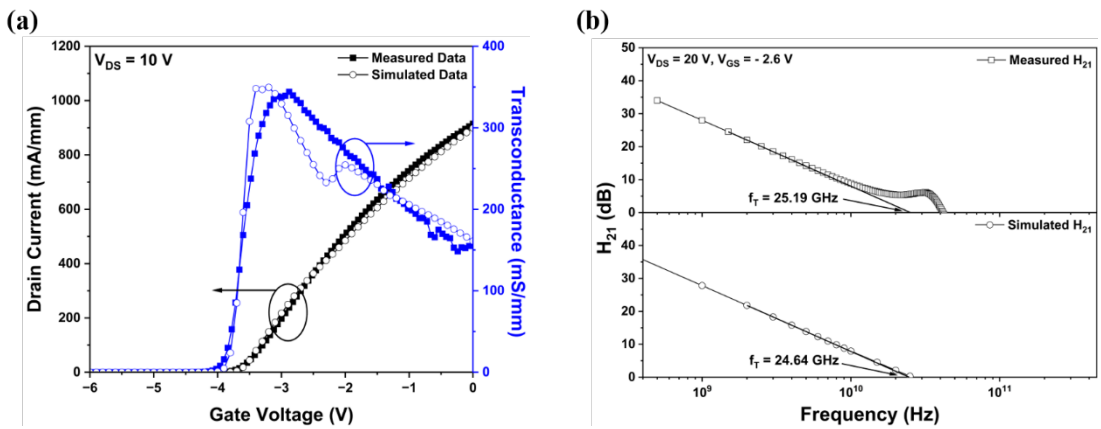


Fig. 2. Comparison between measured and simulated characteristics of the reference Si_3N_4-passivated HEMT: (a) transfer characteristics and (b) current gain as a function of frequency.

3. Comparison of Si_3N_4, HfO_2, and Hybrid Passivation

Three passivation schemes were investigated while keeping the remaining device geometry same. The first was the calibrated reference structure using Si_3N_4 for both passivation layers. In the second structure, HfO_2 replaced Si_3N_4 in the passivation structure. In the third, hybrid passivation, HfO_2 was used only as the first 50-nm passivation layer and Si_3N_4 was retained as the second layer. This comparison isolates the influence of dielectric material and the amount of HfO_2 surrounding the gate and field-plate region.

3.1 DC and Breakdown Characteristics

The HfO_2-containing structures exhibited a small improvement in on-state behavior. The extracted R_{on} values were 4.02 Ω·mm for the conventional Si_3N_4 structure, 3.84 Ω·mm for the HfO_2 structure, and 3.97 Ω·mm for the hybrid structure. In the output characteristics, drain current decreased at higher drain bias because self-heating increased the lattice temperature and enhanced phonon scattering. This effect occurred in all three structures and did not change the overall trend that HfO_2-containing passivation produced slightly higher current.

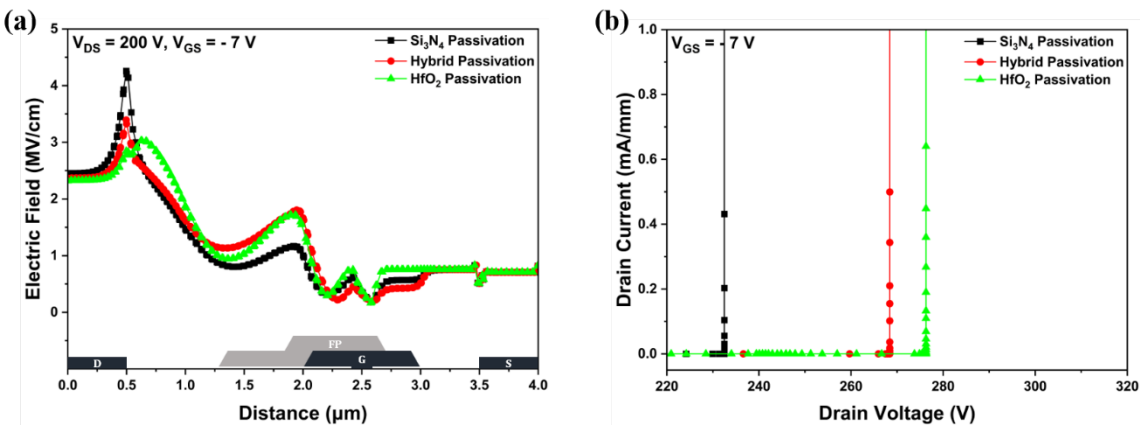


Fig. 3. Comparison of the DC characteristics for different passivation structures: (a) electric-field distribution in the channel and (b) off-state breakdown characteristics.

A more pronounced effect was observed under off-state conditions. At V_{GS} = −7 V and V_{DS} = 200 V, the conventional Si_3N_4 structure exhibited a stronger electric-field peak in the channel as shown in **Fig. 3(a)**. The HfO_2 and hybrid structures reduced and redistributed this peak because the high-k dielectric altered the electrostatic boundary conditions. Since severe electric-field crowding accelerates impact ionization and electron–hole pair generation, a more distributed field profile is beneficial for higher breakdown.

Using a drain-current criterion of 1 mA/mm to define breakdown, V_{BD} was 232.47 V for the Si_3N_4 reference, 276.27 V for the full HfO_2 structure, and 268.41 V for the hybrid structure as shown in **Fig. 3(b)**. Thus, full HfO_2 passivation improved V_{BD} by approximately 19%, while the hybrid structure still provided an improvement of about 16%. The HfO_2 structure also showed the lowest off-state drain leakage. The combination of wider bandgap and modified electrostatic confinement reduces leakage and improves the effective confinement of the channel charge.

3.2 RF Characteristics and DC–RF Trade-off

The principal disadvantage of full HfO_2 passivation was the increase in parasitic capacitance. At V_{DS} = 20 V and V_{GS} = −2.6 V, the full HfO_2 structure showed the highest C_{gs} and C_{gd} because HfO_2 has a much larger dielectric constant than Si_3N_4. The hybrid structure, which uses less HfO_2, showed intermediate capacitance values. This trend directly affected the RF characteristics.

The simulated f_T values were 24.64, 10.17, and 20.50 GHz for the Si_3N_4, HfO_2, and hybrid structures, respectively. Relative to the reference, full HfO_2 reduced f_T by 59%, whereas the hybrid structure limited the reduction to 17%. The corresponding f_{max} values were 110.27, 48.72, and 88.53 GHz. These results

demonstrate that maximizing the fraction of HfO_2 is not optimal for a device intended to operate at high frequency. Although full HfO_2 gives the highest V_{BD}, the accompanying capacitance penalty substantially reduces RF performance.

4. Optimization of the Hybrid Passivation Structure

Because the hybrid structure retained most of the breakdown improvement while recovering a large portion of the RF performance, it was selected for further optimization. The first HfO_2 passivation layer was fixed at 50 nm, and the thickness of the second Si_3N_4 passivation was varied among 150, 250, 350, and 450 nm. Changing this thickness modifies both the distance between the field plate and the channel and the effective electrode separation that determines parasitic capacitance.

4.1 Breakdown Dependence on Second-Passivation Thickness

The electric-field distributions for the four hybrid structures were similar overall, and the peak field did not vary as strongly as it did when the dielectric material itself was changed. The resulting V_{BD} values were 262.00, 268.41, 267.57, and 262.30 V for second-passivation thicknesses of 150, 250, 350, and 450 nm, respectively. The 250- and 350-nm structures therefore provided the highest breakdown voltages.

Two competing mechanisms explain this behavior. Increasing the passivation thickness can help distribute the electric field and reduce the influence of the strong field adjacent to the gate. However, if the second passivation becomes too thick, the source-connected field plate moves farther from the channel and its ability to control the channel electric field weakens. The balance between these effects produces an intermediate optimum rather than a monotonic increase in V_{BD}.

4.2 Frequency Dependence on Second-Passivation Thickness

The RF characteristics showed a clearer thickness dependence as summarized in **Fig. 4**. The 150-nm structure had the largest C_{gs} because the effective gate-to-source electrode separation was the smallest. Increasing the second Si_3N_4 thickness reduced C_{gs} and consequently increased f_T. The simulated f_T values were 17.92, 20.50, 22.64, and 24.97 GHz for 150, 250, 350, and 450 nm, respectively. The corresponding f_{max} values were 78.50, 88.53, 91.47, and 106.39 GHz.

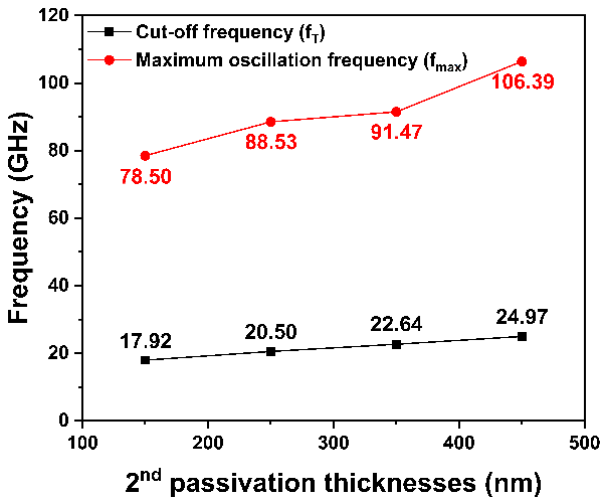


Fig. 4. Simulated cut-off frequency (f_T) and maximum oscillation frequency (f_{max}) of the hybrid HfO_2/Si_3N_4 HEMT as a function of second-passivation thickness.

Although the 450-nm structure nearly recovered the f_T of the conventional Si_3N_4 reference, its V_{BD} decreased to 262.30 V and its lateral figure-of-merit was reduced. The 350-nm structure offered a more balanced solution: V_{BD} remained high at 267.57 V while f_T increased to 22.64 GHz and f_{max} to 91.47 GHz. This result illustrates that passivation thickness should be optimized based on the target application rather than by

considering either breakdown or RF behavior alone. For comparison, LFOM (Standard lateral figure-of-merit) is defined as V_{BD}^2/R_{on}, and JFOM (Johnson's figure-of-merit) as $f_T \times V_{BD}$. The key DC and RF performance metrics are summarized in **Table 1**.

Table 1. Summary of representative simulated device characteristics.

Structure	1st/2nd passivation thickness	R_{on}(Ω·mm)	V_{BD} (V)	f_T (GHz)	f_{max} (GHz)	LFOM (MW/mm)	JFOM (THz·V)
Si_3N_4	50/250 nm	4.02	232.47	24.64	110.27	13.44	5.73
HfO_2	50/250 nm	3.84	276.27	10.17	48.72	19.93	2.81
Hybrid	50/250 nm	3.97	268.41	20.50	88.53	18.15	5.50
Hybrid	50/350 nm	4.16	267.57	22.64	91.47	17.21	6.06

The figure-of-merit comparison emphasizes the same trade-off. Full HfO_2 passivation increased LFOM by approximately 48% relative to Si_3N_4 because of the higher V_{BD} and slightly lower R_{on}, but JFOM decreased by about 51% because f_T was strongly degraded. In contrast, the 350-nm hybrid structure achieved an LFOM of 17.21 MW/mm while increasing JFOM from 5.73 to 6.06 THz·V. Hence, the optimized hybrid design improves the combined voltage–frequency metric while retaining a substantial improvement in lateral power performance.

5. Conclusion

This work examined the influence of HfO_2-based passivation on both DC and RF characteristics of AlGaN/GaN HEMTs using a calibrated TCAD simulation. The reference model reproduced the measured drain current, transconductance, threshold voltage, and cut-off frequency of a fabricated Si_3N_4-passivated HEMT with errors of only a few percent, providing confidence in the comparative simulations.

Replacing Si_3N_4 with HfO_2 significantly improved off-state characteristics. The higher dielectric constant redistributed the electric field in the channel and increased V_{BD} from 232.47 to 276.27 V, while also reducing leakage. However, full HfO_2 passivation increased C_{gs} and C_{gd} and reduced f_T from 24.64 to 10.17 GHz. This confirms a fundamental DC–RF trade-off associated with high-k passivation.

The HfO_2/Si_3N_4 hybrid structure provided a practical means of controlling this trade-off. By restricting HfO_2 to the first passivation layer and optimizing the second Si_3N_4 thickness, high breakdown capability could be retained without the severe RF degradation of the full-HfO_2 structure. Among the evaluated cases, the 50-nm HfO_2/350-nm Si_3N_4 structure showed the best performance achieving V_{BD} = 267.57 V, f_T = 22.64 GHz, f_{max} = 91.47 GHz, LFOM = 17.21 MW/mm, and JFOM = 6.06 THz·V.

The results show that passivation engineering is not only a surface-reliability issue but also an electrostatic design variable that directly affects field distribution, capacitance, and overall device performance. Hybrid high-k passivation with carefully selected dielectric thickness is therefore a promising approach for GaN HEMTs targeting applications that require both high-voltage robustness and high-frequency operation. Future work may extend this concept to experimentally fabricated hybrid-passivation devices and investigate interface-state density, dynamic R_{on}, current collapse, and long-term reliability.

References

1. Zhang, N.Q., Keller, S., Parish, G., Heikman, S., DenBaars, S.P., Mishra, U.K.: High breakdown GaN HEMT with overlapping gate structure. IEEE Electron Device Lett. 21, 421–423 (2000).
2. Meneghini, M., et al.: Reliability and failure analysis in power GaN-HEMTs: An overview. In: IEEE International Reliability Physics Symposium (IRPS) (2017).
3. Mishra, U.K., Shen, L., Kazior, T.E., Wu, Y.-F.: AlGaN/GaN HEMTs—an overview of device operation and applications. Proc. IEEE 90, 1022–1031 (2002).
4. del Alamo, J.A., Joh, J.: GaN HEMT reliability. Microelectron. Reliab. 49, 1200–1206 (2009).
5. Kim, J.G., Cho, C., Kim, E., Hwang, J.S., Park, K.H., Lee, J.H.: High breakdown voltage and low-current dispersion in AlGaN/GaN HEMTs with high-quality AlN buffer layer. IEEE Trans. Electron Devices 68, 1513–1517 (2021).

Design and Architecture Division

Aims

This session presents academic proceedings and practical projects from the Design and Architecture Division. Grounded in a deep understanding of historical contexts, it explores sustainable urban futures through the preservation and regeneration of the built environment. The featured research and real-world case studies demonstrate the dynamic synergy between academic theory and practical implementation.

Scope

- **City and Urban History**
  - Historical evolution of urban morphology
  - Documentation and conservation of historical urban landscapes
  - Socio-cultural impacts of urbanization

- **Management of Architectural Heritage**
  - Practical strategies for conservation and restoration
  - Contemporary utilization and value creation of heritage
  - Advanced heritage management and disaster risk reduction

- **Urban Design**
  - Design projects for public spaces and streetscapes
  - Implementations of placemaking and urban aesthetics
  - Pedestrian-friendly and eco-friendly urban spaces

- **Adaptive Reuse**
  - Adaptive reuse and renovation of existing buildings
  - Sustainable retrofitting of aging structures
  - Extending the lifecycle of the existing built environment

- **Urban Planning and Governance**
  - Master planning and land-use strategies for sustainability
  - Urban governance and citizen participation
  - Policy support and institutional improvements for urban regeneration

Transformation of Berlin's Urban Structure through Long-Distance Railway Network Evolution: A Diachronic Analysis of Major Stations from a Spatial Metabolism Perspective

Sungdong Lee

VeKNI e.V. Berlin, Germany

sungdonglee03@gmail.com

Abstract. Historically, Berlin's urban and spatial transformations have been deeply intertwined with the development and restructuring of its long-distance railway network. However, conventional transportation research often adopts a technical or logistics-centric perspective, failing to fully capture how railway infrastructure interacts with broader dynamic urban evolution—specifically, urban metabolism. This paper investigates the impact of long-distance railway network developments on Berlin's spatial restructuring across three key historical epochs: the Kingdom of Prussia, the German Empire, and the Federal Republic of Germany (Reunified Berlin). Employing a matrix analytical framework that correlates historical political/spatial boundaries with architectural adaptations, this study performs a diachronic comparative analysis of three iconic railway stations: Potsdamer Bahnhof, Lehrter Bahnhof (now Berlin Hauptbahnhof), and Schlesischer Bahnhof (now Ostbahnhof). The findings reveal that these stations underwent significant metabolic transitions, evolving from single-direction peripheral gateways and geopolitically severed terminals into three-dimensional, polycentric, and multi-functional urban nodes. Ultimately, this research demonstrates that long-distance railway networks operated not merely as physical transport corridors, but as primary metabolic organs driving spatial reintegration, functional diversification, and polycentric urban growth in Berlin.

Keywords: Urban Metabolism, Railway Infrastructure, Berlin Urban History, Potsdamer Bahnhof, Berlin Hauptbahnhof, Spatial Restructuring

1. Introduction

Berlin presents a unique urban typology shaped by abrupt geopolitical shifts, structural divisions, and rapid territorial expansions. Since the mid-19th century, the development of long-distance railway networks has played a fundamental role in defining the physical form, economic dynamics, and spatial hierarchy of the city. Despite the extensive literature on Berlin's urban evolution, existing transportation studies often adopt a deterministic or engineering-centric perspective, prioritizing technical capacities, route efficiency, and flow optimization. Consequently, these approaches frequently overlook the dynamic, systemic interplay between transportation infrastructure and the broader urban fabric—a process best conceptualized through the lens of urban metabolism.

Urban metabolism interprets the city as a dynamic, living organism that absorbs, transforms, and redistributes physical flows, human capital, and spatial functions. Infrastructure networks, particularly long-distance railways, operate as the circulation systems of this organism. When railway networks undergo physical expansion, spatial interruption, or structural realignment, they trigger corresponding shifts in urban land-use patterns, civic centrality, and spatial accessibility.

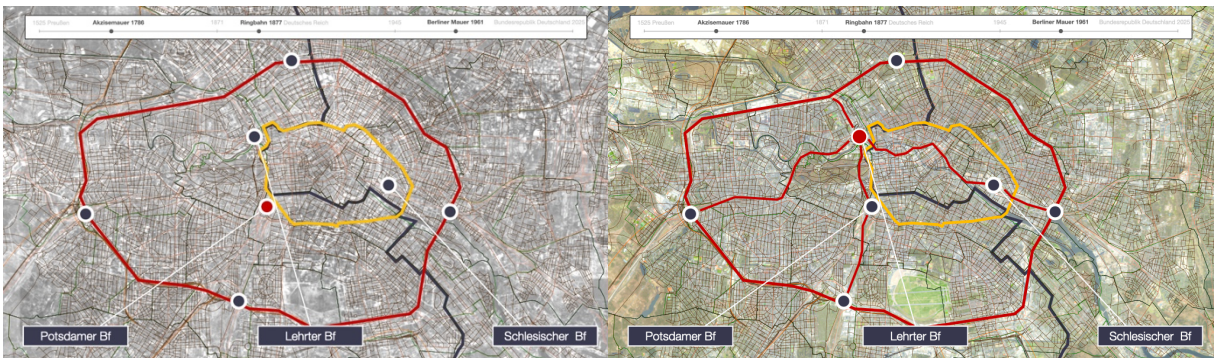
To systematically analyze this process, this study investigates how long-distance railway developments restructured Berlin's spatial organization across three pivotal historical eras:

1. The Prussian Kingdom Era (mid-19th century): Initial network formation along urban peripheries and fiscal boundaries.
2. The German Empire Era (late 19th to early 20th century): Consolidation of radial and orbital rail lines (Ringbahn) amidst rapid industrialization.
3. The Reunified German Era (post-1990 to present): Network reintegration, structural centralities, and modern urban regeneration.

The research focuses on three major railway station sites that served as strategic nodes throughout these epochs: Potsdamer Bahnhof, Lehrter Bahnhof (now Berlin Hauptbahnhof), and Schlesischer Bahnhof (now Ostbahnhof). By employing a timeline-boundary matrix that aligns historical political thresholds (e.g., the Akzisemauer, the Ringbahn, and the Berliner Mauer) with spatial functional transitions, this paper elucidates the structural relationship between transport infrastructure and urban metabolic evolution.

2. Case Studies and Spatial Transformations

2.1 Potsdamer Bahnhof: Transition from Urban Gateway to Mixed-Use Civic Hub



Established in 1838 as the terminus of the Berlin-Potsdamer Eisenbahn under King Friedrich Wilhelm III, Potsdamer Bahnhof was sited immediately outside the southwestern customs wall (Akzisemauer) near the Potsdamer Tor. Initially operating as an peripheral terminus, the station functioned as a unidirectional conduit, funneling external passenger and goods flows into the historic city core while reinforcing the territorial boundary of the customs wall. As industrialization accelerated, the area surrounding Potsdamer Bahnhof rapidly evolved into a dense commercial and transit node.

During the early 20th century, the station underwent major technological and spatial adaptations, including track subterraneanization and the opening of the S-Bahn Potsdamer Platz station in 1939. This established Potsdamer Platz as one of Europe’s most intensely utilized urban intersections. However, extensive wartime destruction followed by the cold war division—symbolized by the construction of the Berlin Wall in 1961—rendered the site a vast urban wasteland (Brachland).

Following German reunification, the site underwent a dramatic metabolic transformation. Through mega-scale urban regeneration projects in the 1990s and the continuous development of public open spaces such as the Park am Gleisdreieck (completed around 2023), the former railway enclave was completely reconfigured. Synthesizing its diachronic trajectory, Potsdamer Bahnhof transformed from a peripheral, single-purpose gateway terminus into a highly integrated, multi-layered urban center and civic public space, exemplifying how an infrastructure node can shift from marking an urban boundary to synthesizing civic life.

2.2 Lehrter Bahnhof: From Interrupted Terminal to Three-Dimensional Central Interchange

Opened in 1868, Lehrter Bahnhof served as the terminus for the railway line connecting Berlin with Hanover and northwestern Germany. Its strategic position was further enhanced by its proximity to the Berliner Ringbahn (1877), positioning the northern Spree curve as a critical logistics and passenger junction. However, the political division of Berlin after World War II and the subsequent erection of the Berlin Wall in 1961 severely disrupted this network. Situated directly along the border sector, Lehrter Bahnhof lost its long-distance operational role, resulting in physical isolation, station demolition, and spatial fragmentation of the surrounding urban fabric.

The reunification of Germany necessitated a comprehensive overhaul of Berlin's disconnected railway infrastructure. In the 1990s, the Senate of Berlin and Deutsche Bahn adopted the Pilzkonzept ("Mushroom Concept"), a master planning framework designed to replace the historic, fragmented terminal system with an integrated, multi-directional network. Under this framework, the former site of Lehrter Bahnhof was selected for the construction of Berlin Hauptbahnhof, opened in 2006.

Architecturally and urbanistically, Berlin Hauptbahnhof represents a technological marvel of spatial synthesis. Featuring a two-level, multi-tiered grid where the East-West elevated line (Stadtbahn) crosses the

North-South subterranean line (Tiergarten tunnel) at right angles, the station converts a former zone of division into an active spatial hinge. The diachronic evolution of Lehrter Bahnhof demonstrates a qualitative leap from a geopolitically severed border node to a three-dimensional, multi-scalar international transit interchange, establishing a unified circulation spine for the metropolis.

2.3 Schlesischer Bahnhof: Evolution from Monocentric Logistics Hub to Polycentric Metabolic Axis

Inaugurated in 1881, Schlesischer Bahnhof (later renamed Hauptbahnhof East, and currently Berlin Ostbahnhof) was designed as the principal eastern terminal connecting Berlin to Silesia, Eastern Europe, and the industrial belts along the Spree. Linked directly with the Berliner Ringbahn, Schlesischer Bahnhof formed the backbone of Berlin's eastern industrial ecosystem. Throughout the late 19th and early 20th centuries, the station operated as a heavy metabolic organ, importing raw materials and agricultural goods while exporting manufactured products.

During the East German (GDR) era, the station retained its prominence as the chief long-distance railway hub for East Berlin. However, the post-reunification urban agenda shifted Berlin away from a traditional monocentric development model toward a polycentric (multinodal) spatial framework. Within this renewed planning context, Ostbahnhof underwent functional diversification. Rather than serving purely as a heavy industrial/freight handling facility, the station area was redeveloped into a vibrant cultural, commercial, and residential district integrated with the East Side Gallery precinct.

The structural trajectory of Schlesischer Bahnhof illustrates a fundamental functional realignment: transitioning from a monocentric industrial and logistics hub into a key polycentric axis within Berlin's distributed urban metabolic network. This shift emphasizes that contemporary transport nodes must accommodate complex, multi-functional land uses beyond traditional transportation services.

3. Conclusions

This study has investigated the diachronic evolution of three pivotal Berlin railway stations—Potsdamer Bahnhof, Lehrter Bahnhof, and Schlesischer Bahnhof—to demonstrate how long-distance transportation networks drive the dynamic restructuring of urban space. By examining these sites through a spatial metabolism framework across Prussian, Imperial, Cold War, and Reunified eras, several critical conclusions emerge:

1. Overcoming Boundary Constraints: Railway infrastructure in Berlin evolved from peripheral boundary markers (Akzisemauer) into powerful drivers of internal spatial integration and urban expansion.
2. Network Synthesis and Multi-dimensionality: The transformation of Lehrter Bahnhof into Berlin Hauptbahnhof via the Pilzkonzept illustrates how a fragmented array of historical terminus stations can be re-engineered into an integrated, three-dimensional interchange grid that heals urban ruptures.
3. Polycentric Regeneration: The shifts observed at Potsdamer Bahnhof and Schlesischer Bahnhof highlight the transition from single-function transport enclaves toward multi-functional, polycentric urban nodes that support civic, economic, and cultural activities.

In summary, the evolution of Berlin's long-distance rail network reflects the continuous adaptation of an urban organism responding to political, technological, and spatial imperative. These findings suggest that contemporary megacity infrastructure planning must look beyond immediate throughput capacity, approaching transport nodes as vital metabolic catalysts that shape long-term spatial resilience and civic vitality.

References

1. Wolman, A.: The metabolism of cities. Scientific American 213(3), 179–190 (1965).
2. Kennedy, C., Cuddihy, J., Engel-Yan, J.: The study of urban metabolism. Journal of Industrial Ecology 11(2), 43–59 (2007).
3. Gottwaldt, A.: Berliner Fernbahnhöfe: Erinnerungen an ihre große Zeit. Alba Publikation, Düsseldorf (1991).
4. Ladd, B.: The Ghosts of Berlin: Confronting German History in the Urban Landscape. University of Chicago Press, Chicago (1997).
5. Balfour, A.: Berlin: The Politics of Order 1737–1989. Rizzoli, New York (1990).
6. von Gerkan, M.: Berlin Hauptbahnhof. Jovis Verlag, Berlin (2006).
7. Bley, P.: 150 Jahre Eisenbahn Berlin–Frankfurt (Oder). Alba Publikation, Düsseldorf (1998).

Polycentricity.
Historical Transformations and Urban Regeneration in Central Berlin

Kwangjin Lee

K LEE ARCHITEKT, Sachtlebenstraße 53, 14165 Berlin, Germany
k.lee.architekt@gmx.de

Abstract. This study stems from the recognition that understanding a city's historical context is essential for creating sustainable urban and architectural spaces amid rapid urban changes. Unlike other historic European capitals, Berlin grew into a relatively young metropolis. As the first stage of this study, we trace how "polycentricity"—the core spatial characteristic of this city—was historically formed and altered.
In the second stage, we conduct a precise analysis of specific district-level regeneration plans and high-rise developments. This analysis focuses on three key monitoring tasks currently unfolding in central Berlin post-reunification: T1 (Spatial Restoration of the Historic Centre), T2 (New Urban Spatial Connection between the Old Town and Alexanderplatz), and T3 (Reconnecting Spatial Flows in City West). Through this, we aim to examine how the recovery of historical texture and modern high-density development evolve into a balanced state within a polycentric network structure.

Keywords: Polycentricity, Urban Space, Berlin, historical transformation

1. Introduction

Modern cities face the challenge of securing sustainability amidst rapid socioeconomic transformations, such as climate change, population migration, and technological innovation. For architectural, technical, and sociological efforts to yield practical results in addressing these challenges, they must be preceded by an understanding of the historical context—specifically, how cities and buildings have been formed and changed over time. This is because a city's unique spatial assets and identity are accumulated through long historical processes, and development that overlooks this threatens urban sustainability. With this awareness, this research was conducted as part of the "City & Urban History" project, which has been carried out since 2024 by the research group "KLEE & YAHN" with support from the Architecture and Urban Planning Division of the Association of Korean Scientists and Engineers in Germany (VEKNI). Rather than forming a single core, Berlin possesses a unique "polycentricity" where multiple centers interact with one another. This study goes beyond a first-stage analysis of historical context to provide an in-depth, structured analysis of the real-time spatial transformations and urban regeneration issues across these centers—categorized into three axes: T1, T2, and T3—to maintain the polycentric balance of modern downtown Berlin.

2. Stages of Historical Transformation in Berlin's Spatial Structure

The medieval core of central Berlin has undergone dramatic spatial structural transformations, passing through three major historical turning points up to the modern era.

18th–19th Century: New Town Development and Spatial Isolation (The Hobrecht Plan Era)

In the 17th century, Berlin's historical center was located around today's Museum Island and its eastern periphery, enclosed by city walls. At that time, the downtown area formed a circular spatial framework connecting Schlossplatz, Cölnischer Rathaus, Petriplatz, and Molkenmarkt. Entering the 18th century, a grid-patterned new town (Friedrichstadt), which included three large squares, was developed west of this historic downtown, separated by the city wall. In the 19th century, a series of urban architectural works by Karl Friedrich Schinkel strongly connected the urban spaces of the new town with the palace spaces. However, this connection failed to effectively integrate with the existing circular structure of the historic downtown. Furthermore, as "Karlschlagsanierung" (a large-scale urban clearance project involving widespread destruction) took place in the late 19th century, the circular core of the old town became spatially isolated, maintaining a meaningful connection only with Alexanderplatz to the east. In contrast, the new town, equipped with a Baroque grid structure and major squares, connected strongly with the outer suburbs and

emerged as Berlin's new central axis. During this period, Peter Joseph Lenné's concept of a "chain of green spaces" and the 1862 "HobrechtPlan" laid the foundation for Berlin's characteristic polycentric structure by forming multiple independent new centers in the peripheral areas.

20th Century: Destruction of War and Distortion by East-West Division

World War II completely devastated central Berlin, and the construction of the Berlin Wall during the Cold War severely deformed and distorted the city's spatial structure. Although East Berlin contained the old historical town, instead of restoring it, the authorities remodeled the area into a massive, automobile-oriented square reflecting socialist ideology. They built the Palace of the Republic (Palast der Republik) on the site of the former city palace and erected a giant open space along with the TV Tower (Fernsehturm) to the east, creating a monumental center. On the other hand, Molkenmarkt and Petriplatz in the southern part of the old town were completely destroyed for road expansion. Simultaneously, West Berlin developed "City West" into a new, independent center by creating a square (Breitscheidplatz) in front of the ruined Kaiser Wilhelm Memorial Church and expanding the Kurfürstendamm avenue. Additionally, they solidified their western foothold by constructing a roundabout called Ernst-Reuter-Platz at the western edge of the Tiergarten.

2.3. Post-1990: Reunification and Spatial Reintegration

The reunification of Germany in 1990 brought the challenge of physically reconnecting a divided urban space. North of the Tiergarten, the new central station (Berlin Hauptbahnhof) was constructed, completely restructuring the railway system, while the ribbon-shaped "Government District" was established to the south to connect East and West. Potsdamer Platz, which was once the boundary of the new town, was redeveloped into a financial and office center filled with high-rise buildings, linking up with West Berlin's Kulturforum. In terms of green infrastructure, Berlin Wall Park (Mauerpark) and the Luisenstadt Canal were restored as linear parks along the former wall line, allowing the central structure to evolve into a polycentric network.

3. Modern Berlin's 3 Urban Regeneration and Monitoring Issues

Having entered a mature phase post-reunification, Berlin is implementing sophisticated urban regeneration and high-rise plans across individual districts to accommodate the conflicting demands of historical identity recovery and polycentric functional enhancement.

Fig. 1. Transformation of urban spatial structure in central Berlin 2025, with Today's Issues T1, T2 and T3

Issue T1: Restoration of Urban Spaces in the Historic Centre

The core targets of T1 are the Molkenmarkt and Petriplatz districts, whose urban tissues were disrupted due to the expansion of automobile roads during the Cold War era. **Molkenmarkt District:** In order to recover the historic medieval core grid of 1652, a development plan (Bebauungsplan) was established in 2016. In the 2021 project competition, proposals such as the "Bernd Albers & Silvia Malcovati" plan and the "OS Arkitekter" plan competed. Currently, the creation of a "New Block" is underway to reinterpret the historic street structure in a modern way. The core goal is to restore medieval density and street facades to a downtown area that had been torn apart by massive roadways. In terms of restoring the historical center's original circular spatial structure, this approach can be evaluated positively. This recovery, paired with the adjacent Nikolaiviertel (Nikolai Quarter), will be understood as a renewed expansion and restoration of Berlin's old town. However, fully reclaiming the central urban area as a pedestrian-oriented space appears highly challenging. Consequently, the critical factor will depend on how effectively the outdoor spaces within each block are provided as meaningful public urban domains. Furthermore, to shape the diverse building facades characteristic of the historic downtown, it is essential to select multiple small-scale private developers rather than a single large-scale one; this can only be achieved through close cooperation between the private sector and the state government. **Petriplatz District:** Based on the historical placeness of the Cölln Fish Market (CöllnischerFischmarkt, 1758), the construction of three major cultural and religious centers is either in its final stages or currently underway. Integrating the three buildings through a colonnade via the district unit plan (Bebauungsplan) is highly meaningful, as it naturally extends the urban space from the Nikolai Quarter across the river. In this context, Hotel Petriplatz reflects the spatial characteristics of the historical urban texture by reinterpreting the volume and facade composition of the former Cölln Town Hall (Cöllnischer Rathaus). Following this flow, the newly established House of One will replace the former Petri Church (Petrikirche), which historically featured an introverted urban space. This urban composition is then completed by the Petri Berlin Archaeological Centre, through which the once-destroyed urban spatial identity will be newly formed. Moving forward, a critical monitoring task will be observing how these urban spaces connect back to Museum Island (Museumsinsel) as part of the circular spatial structure, and how they link westward to Spittelmarkt in Friedrichstadt.

Issue T2: New Urban Spatial Connection between the Historic Centre and Alexanderplatz

T2 is a project aimed at healing the pedestrian and visual disconnect between the old town (Historic Centre) and the commercial center of Alexanderplatz by readjusting spaces that used to be giant socialist avenues in East Berlin during the division. The continuous five-story skyline spanning across central Berlin has long characterized the city’s historic urban profile. While maintaining this defining feature, three specific high-rise development zones have been urbanistically designated to accommodate modern urban expansion, one of which is Alexanderplatz. Since the 17th century, Alexanderplatz evolved in close connection with the old town's circular spatial structure. However, through the construction of the S-Bahn station in the 19th century, followed by the mid-to-late 20th-century socialist restructuring and automobile-oriented transformation under East Germany, the square became functionally complex yet spatially isolated from its surroundings.In this context, the recent completion of the Humboldt Forum, the progressive high-rise clusters at Alexanderplatz since the 1990s, and the transformation of the Rathaus- and Marx-Engels-Forum into an urban park will collectively serve to strongly reconnect Alexanderplatz with Museum Island and the Spree River. Paradoxically, the massive open spaces created during the East German era now provide the vital foundation for delivering ample urban nature to these high-rise clusters, ultimately linking them to the Spree River—the city’s central ecological axis. Due to the current economic stagnation across Germany, completing these high-rise clusters will still require significant time and encounter various modifications. Furthermore, the emergence of these high-rise complexes is anticipated to heavily impact surrounding historic urban tissues, particularly the commercial ecosystem around Hackescher Markt. Finding an appropriate equilibrium with the existing neighboring urban spaces will remain a critical challenge moving forward.

Issue T3: Reconnection of Urban Spatial Flow in City West

T3 is a master plan designed to correct the disrupted pedestrian flows of City West—the former center of West Berlin—and to develop the area surrounding the Zoologischer Garten transport hub into a high-density, three-dimensional mixed-use zone. **Hertzallee / Hardenbergplatz Framework Plan 2022:** The basic

framework of City West established by the 1862 Hobrecht Plan had its northward flow cut off by railway junctions and other barriers. In 2022, the City of Berlin established a new framework plan to completely remodel Hardenbergplatz in front of the Zoo Station into a pedestrian-friendly space, while designating the Hertzallee area—previously underutilized railway land—as a high-rise mixed-use zone.**City West High-Rise Structure:** In addition to already constructed buildings like Zoofenster (120m), Upper West (120m), and the building behind the Kaiser Wilhelm Memorial Church (KWGK, 71m), a three-dimensional land-use plan has been finalized and is being executed. This plan places a cluster of new high-rise building complexes (New Tower Complex Area) ranging from 75m, 80m, 100m, up to a maximum of 110m in the core district north of the railway station. Through this, Lenné's urban spatial flow—which had been severed and weakened by railway lines and 20th-century developments—is being revitalized once again.

4. Conclusions

This study analyzed how Berlin's central area evolved away from a single core into a centrifugal, flexible, polycentric network city (PolycentriCity), and examined three modern regeneration initiatives (T1, T2, T3) designed to sustain this structure. In synthesis, Berlin's polycentricity is not a stagnant historical fragment; rather, it is a dynamic structure that continuously reorganizes itself while absorbing modern development pressures. Issue T1 (Old Town) represents an urban spatial transformation that meticulously and archaeologically restores medieval plots and street grids to recover historical identity. Issue T2 (Alexanderplatz) is an attempt to shape a new central urban space for Berlin through high-rise building clusters and large-scale urban green spaces. Issue T3 (City West) likewise pursues the formation of high-rise building clusters, but it is understood as an effort to restore the urban spatial quality of the 19th century through these additions. These diverse urban spatial endeavors combine organically with the Tiergarten—the massive natural buffer zone inside the city center—and linear green axes (Mauerpark, Luisenstadt Canal) to neutralize urban overcrowding and mitigate spatial conflicts. Ultimately, by integrating historical restoration (T1), large-scale capital-driven high-density development (T2, T3), and the reconnection of environmental infrastructure utilizing the heritage of the Berlin Wall into a polycentric structure, Berlin presents the most flexible and sustainable urban space model for a modern metropolis to follow.

References

1. HistoMap Berlin, https://histomapberlin.de/de/index.html, last accessed 2025/05/09.
2. Geoportal Berlin, https://gdi.berlin.de/viewer/main/, last accessed 2025/05/09.
3. Berliner Stadtplansammlung, https://www.berliner-stadtplansammlung.de, last accessed 2025/05/09.
4. Architekten- und Ingenieurverein zu Berlin-Brandenburg e.V.: Unvollendete Metropole Band 1, DOM publishers, Berlin (2020)
5. Lampugnani, V.M.: Die Stadt von der Neuzeit bis zum 19.Jahrhundert, Verlag Klaus Wagenbach, Berlin (2017)
6. Schlug, A.: 5 Plätze – 1 Name. Der Berliner Alexanderplatz, Vergangenheits Verlag (2008)
7. Senators für Stadtentwicklung und Umweltschutz, Berlin/West: Die Plätze am Kurfürstendamm, Bestandsaufnahme und Empfehlungen zu den Plätzen und platzräumlichen Situationen am Kurfürstendamm, Schriften des Senators für Stadtentwicklung und Umweltschutz, Berlin, Konopka (1985)
8. Bodenschatz, H.: Städtebau in Berlin. Schreckbild und Vorbild für Europa, DOM publishers, Berlin (2013)
9. Bodenschatz, H., Tubbesing, M., Zohlen, G.:100 Jahre (Groß-) Berlin und seine Zentren, Wasmuth &Zohlen Verlag, Berlin (2019)
10. Bodenschatz, H., Flierl, T.: Berlin plant. Plädoyer für ein Planwerk Innenstadt Berlin 2.0, Verlag Theater der Zeit, Berlin (2010)

A Study on the Urban Spatial Transformation of Berlin's Embassy and Government Districts in Response to Political and Social Changes

Inho Woo[1]

[1] Schultes Frank Architekten, Lützowplatz 7, 10785 Berlin, Germany
wih.berlin@gmail.com

Abstract. This study analyzes the impact of political and social regime changes from the 19th century to the present on the formation, destruction, and reorganization of Berlin's key urban spaces: the government district (Wilhelmstraße and Spreebogen) and the embassy district (Tiergarten Süd and Alsenviertel), focusing on the changes in spatial characteristics by region. During the Kingdom of Prussia and the German Empire, Berlin formed high-density, closed-block spaces concentrated with elite power; however, through the uniform urban planning of the Nazi regime, the physical destruction of World War II, and the East-West division during the Cold War, it experienced the denial and dissolution of these power spaces. Following the reunification in 1990, Berlin was reborn as a center of a new democracy that emphasizes transparency and citizen accessibility through low-density, open spatial organizations such as the 'Band des Bundes' (Ribbon of the Federal). By tracing the changes in urban form, spatial organization, function, political context, and social meaning that each region underwent over time, this paper establishes that urban space is not merely a physical background, but a medium that embodies a nation's regime changes and the zeitgeist of the era.

Keywords: Berlin, Urban Space, Government District, Embassy District, Wilhelmstraße, Tiergarten Süd, Alsenviertel, Band des Bundes.

1. Introduction

Urban space is a material manifestation of the political ideology and social demands of the nation occupying it. Over the past 200 years, Berlin has been a global center that experienced dramatic regime changes, including the Weimar Republic, the Nazi regime, the Cold War division, and reunification. These shifts in power and regime transitions constantly altered the location and form of the government district, the core space of power, and the embassy district, the arena for diplomatic negotiations. This study traces the spatial transformation of three key districts—Alsenviertel (Spreebogen), Tiergarten Süd, and Wilhelmstraße—according to historical flows, and in-depth analyzes how their urban forms, spatial organizations, and corresponding political and social meanings were reconstructed.

2. Historical Analysis and Discussion

Prussian and Imperial Era: Formation of Closed Elite Power Spaces

From the mid-19th century to the period before World War I, Berlin's core districts were organized to concentrate imperial power and flaunt the authority of the elite class.

The urban form of Wilhelmstraße was an authoritative street densely lined with aristocratic palaces (Palais) and government buildings. In terms of spatial organization, it took the form of closed blocks where state power was concentrated, and its main function was to serve as the center of state power execution and diplomatic/administrative affairs. Within the political context of being the heart of state power and diplomacy, this street carried the social meaning of the absolute concentration of elite power among a minority ruling class.

Alsenviertel, the diplomatic space at the time, exhibited a similar pattern. Located in the Spreebogen near the Reichstag, the urban form of this area was a high-density block, and its spatial organization was closed off from the outside and internally oriented. Functioning primarily as the residential area for diplomats and embassies of Nordic countries such as Finland, Norway, and Denmark, this place had a political context that contributed to the concentration of imperial power and formed the social meaning of an exclusive city strictly for the elite and diplomats. Tiergarten Süd, located in the west of the city center, also took the urban form of high-density blocks as an upscale villa neighborhood for the upper class in the pre-war era. This area possessed a spatial organization of closed block interiors centered on aristocrats and diplomats, and it

performed a hybrid function mixing residence, diplomacy, and culture. This implies that, under the political context of being an upper-class diplomatic and cultural district during the Imperial and Weimar periods, this area held the social meaning of a 'cultural annex' supporting state power beyond being a mere residential neighborhood.

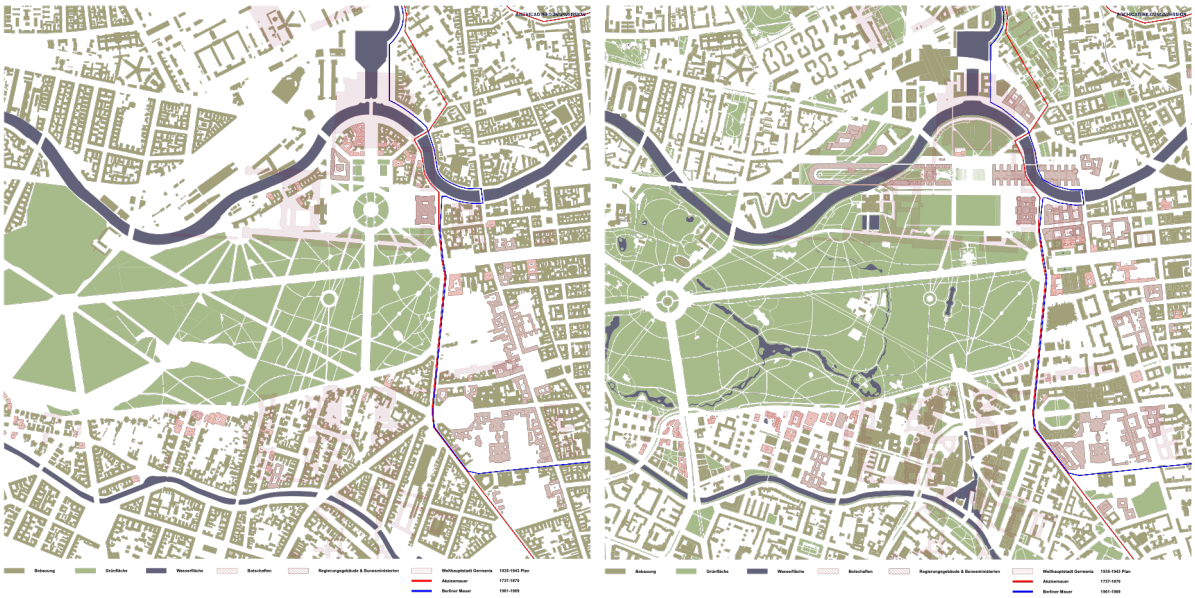


Fig. 1. Berlin in the 1910s and Today: Locations of Government Offices and Embassies in the Alsenviertel, Wilhelmstraße, and Tiergarten-Süd Areas, and the National Socialists' 'Welthauptstadt Germania' Plan

Urban Planning of the Nazi Regime and the Destruction of World War II

The Nazi regime, which took power in 1933, promoted an ambitious urban plan to remodel Berlin into the 'World Capital Germania,' destroying and reorganizing the existing spatial order. The Nazis demolished the Nordic embassies in Alsenviertel and forced their relocation to build the colossal 'Große Halle' (Great Hall) in the Spreebogen area. As a result, the organic diplomatic function of Alsenviertel collapsed, and the embassies of various countries were forced to move to the western part of Tiergarten. Wilhelmstraße was also overlaid with authoritative architecture symbolizing totalitarianism, such as the massive New Reich Chancellery and the Reichsluftfahrtministerium (Ministry of Aviation), maximizing the display of power. The subsequent outbreak of World War II thoroughly destroyed the physical framework of central Berlin, including the magnificent villas and historic embassy buildings in Tiergarten Süd.

Cold War Divided City: Denial of Power and Spatial Disconnection

As post-war Germany was divided into East and West, each region underwent different ideological spatial reorganizations. When West Germany designated Bonn as its provisional capital in 1949, Western diplomatic missions relocated, causing Berlin's core districts to decline or transform into propaganda arenas for the new regimes.

Wilhelmstraße, incorporated into East Germany, faced a complete spatial subversion. Renamed 'Otto-Grotewohl-Straße' in 1964 after the first Prime Minister, the new urban form of this street changed into a uniform structure centered on prefabricated panel apartments (Plattenbau). The formerly closed blocks were depoliticized and replaced by an open-space-oriented spatial organization, and its function shifted from an administrative center to a hub for the daily residential life of citizens. This carried the political context of creating a socialist living space and denying the symbols of past power, inscribing a powerful social meaning of abolishing elitism and pursuing equality and the masses into the space.

Meanwhile, Tiergarten Süd, left in West Berlin, developed in a completely different way with the creation of the 'Kulturforum' upon the ruins. With the introduction of modern cultural facilities such as the Berlin

Philharmonie designed by Hans Scharoun, this area was utilized as a base to transplant the freedom and cultural identity of West Berlin during the division era. The area around the former Reichstag and Spreebogen, where Alsenviertel was located, degenerated into a 'Death Strip' penetrated by the Berlin Wall, and the urban railway network (S-Bahn) crossing East and West was severed, resulting in 'Ghost Stations' where trains passed without stopping.

Post-Reunification: Spatial Reorganization Towards Openness and Transparency

With the reunification of Germany in 1990 and the return of the capital to Berlin, the disconnected spaces of the past were reconstructed into symbols of democracy and integration.

The Spreebogen area, where the Wall previously passed, was chosen as the new government district. The 'Band des Bundes' (Ribbon of the Federal), designed by Axel Schultes and Charlotte Frank, arranged the Chancellery and the Paul-Löbe-Haus in a straight line, embodying a gesture of integration that symbolically crosses the river that once divided East and West Berlin. As seen in the glass dome of the Reichstag designed by Norman Foster, it aims for a political context of transparency and openness to the public. This holds the social meaning of maximizing citizen accessibility to reconstruct the democratic national identity of unified Germany.

Tiergarten Süd also revived as a complex space where culture and diplomacy converge. Its urban form was filled with low-density pavilion-type buildings instead of the previously dense villas, showing a spatial organization where green spaces and buildings are leisurely dispersed. This place performs a hybrid function that blends a diplomatic space concentrated with various embassies and a cultural space of the Kulturforum (Philharmonie, museums, etc.). Within the political context linking unified Germany and the international community, Tiergarten Süd expresses an internationalist identity and has acquired a symbolic social meaning representing the global city of Berlin.

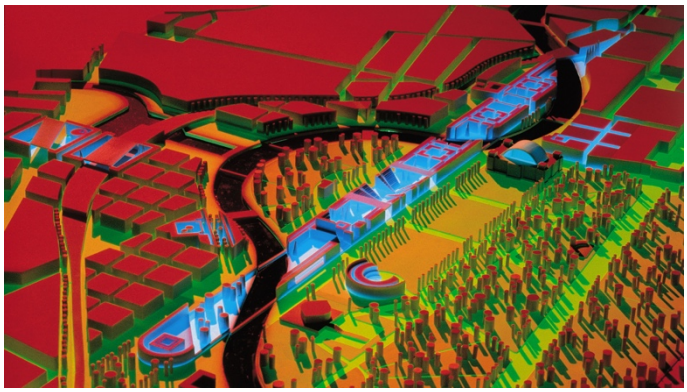


Fig. 2. Winning Entry for the Spreebogen Urban Design Ideas Competition: Architectural Model of the Band des Bundes bySchultes Frank Architekten

Wilhelmstraße evolved anew in a state where past power and East German daily life are intermingled. Its current urban form is a coexistence of East German-era apartments, modern government buildings, and a reorganized administrative axis. In terms of spatial organization, it takes a three-dimensional form where state power is concentrated again but infused with transparent publicness. By housing the Federal Ministry of Finance (former Ministry of Aviation) and the Federal Ministry of Labour and Social Affairs (part of the former Ministry of Propaganda), it restored its function as the center of federal government administration; however, by recycling buildings with historic stigmas rather than destroying them, it created a political context of a transparent democratic administrative center. Consequently, Wilhelmstraße acquired a social meaning where the authority of state administration, civic daily life, and democratic openness intersect.

3. Conclusions

Berlin's government and embassy districts are intersections of urban and political history that physically record the massive regime changes and political ideological shifts from Prussian imperialism, Nazi totalitarianism, and the Cold War division, to democracy and peaceful reunification. The trajectories of each region derived through this study are as follows. Alsenviertel (Spreebogen) evolved from a closed elite space into a transparent riverfront axis (Band des Bundes) open to citizens, and Tiergarten Süd was reborn from a cultural annex of the upper class into the premier diplomatic and cultural destination of the global city of

Berlin, where dispersed green spaces and buildings harmonize. Wilhelmstraße was reorganized from the flaunting of authoritative power through the egalitarian apartment complexes of East Germany (denial of power) into a democratic administrative center aiming for transparent publicness while preserving the memories of the past. This clearly proves that unified Germany successfully sublimated urban space into a social medium that embodies a new zeitgeist (openness, transpare https://maps.google.com/ncy, internationalization) by reinterpreting the fragmented urban forms and spatial organizations, rather than simply erasing the traces of the past.

References

1. Author, Helmut Zschocke.: Geheimnisvolles Alsenviertel am Bundeskanzleramt. Peter Lang, Frankfurt am Main (2017).
2. Author, Axel Schultes., Author, Charlotte Frank.: Kanzleramt Chancellery Berlin. Edition Axel Menges, Stuttgart (2002).
3. Julius Straube.: Übersichtsplan von Berlin in 44 Blättern (1910). Scale 1:4000. Berlin: Geographisches Institut und Landkartenverlag. Original from Landesarchiv Berlin; digitized on Wikimedia Commons, https://de.wikipedia.org/wiki/Datei:Straube-Plan_Berlin_1910.jpg.
4. Landesarchiv Berlin, and Berliner Hochschule für Technik.: HistoMap Berlin: Online Portal for Georeferenced Historical Maps of Berlin (1970–1990). Accessed August 9, 2026. https://histomapberlin.de/en/.

The Formation of Cologne's Historic City Walls, the Green Belt, and the Contemporary Urban Structure

Soonkyo Jeong[1]

[1] STEVES & BORSUM ARCHITEKTEN PartGmbB, Bayenthalgürtel 68, 50968 Köln, Germany
archi4ne1@gmail.com

Abstract. Cologne is one of Germany's most historically significant cities, preserving within its urban fabric traces of successive historical periods from antiquity to the present. Although the city suffered extensive destruction during the Second World War and its post-war reconstruction has often been critically assessed, its urban structure continues to reveal the overlapping layers of Roman, medieval, and modern development. Among these, the Green Belt, established through the transformation of former fortifications, has become a defining element of the city's contemporary identity.

This paper investigates the historical evolution of Cologne's fortification systems and their influence on the present urban structure through cartographic analysis and graphic representation. Historical maps, urban morphology, surviving architectural fragments, and the Green Belt are systematically documented and reinterpreted through a series of mappings and diagrams. By visualizing the spatial continuity and transformation of former defensive structures, the study reveals how historical urban elements remain embedded within the contemporary city and continue to shape Cologne's spatial identity.

Keywords: Cologne, City Wall, Fort, Fortification, Tower, Gate, Green Belt.

1. Introduction

During the nineteenth century, Cologne underwent profound urban transformations. As the historic city center became increasingly constrained by its fortifications, the medieval city walls were demolished, allowing the urban core to expand beyond its former boundaries. Under the direction of the urban planner Hermann Josef Stübben, the district of Cologne Neustadt was subsequently developed as part of this large-scale urban expansion. The concept of the Inner Green Belt originated from Fritz Schumacher's plan of 1919. His initial proposal envisioned a sequence of parks enclosed by surrounding urban blocks. However, subsequent planning revisions transformed this concept into a more open and continuous arrangement of green spaces. The Inner Green Belt emerged from the broader ambition to establish a large-scale urban park encircling the city of Cologne. Although this vision was initiated, it was never fully realized. As a result, the present-day Inner Green Belt exists as a fragmented sequence of parks, open spaces, and vacant lots, with several sections occupied by commercial and industrial uses rather than functioning as a coherent urban park. The challenge of reconnecting these fragmented elements into a unified whole has been repeatedly addressed since the Albert Speer Master Plan and remains a central objective of numerous contemporary planning projects. Future planning initiatives therefore require a comprehensive understanding of the relationship between the Green Belt, the surrounding urban structure, and its historical context. This study aims to contribute to that understanding through the analysis and cartographic representation of Cologne's historical urban development.

2. Urban expansion and structural changes according to historical trends

The contemporary urban structure of Cologne retains the visible traces of its historical development. Former defensive boundaries have not simply disappeared but have been transformed and incorporated into the present-day urban fabric. The course of the Roman city wall remains legible in parts of the existing street network, while the line of the medieval fortifications has been reinterpreted as the ring roads that today form the principal circulation system around the historic centre. Likewise, sections of the railway infrastructure occupy corridors formerly associated with military fortifications. The former military grounds of the Prussian fortress system were subsequently converted into public green spaces, providing the spatial foundation for the development of Cologne's Green Belt.

Figure 1 illustrates the successive transformation of Cologne's urban structure, from the Roman city wall and the medieval fortifications to the Prussian fortress system and, ultimately, the formation of the contemporary Green Belt. Rather than depicting a simple chronological sequence, the diagram reveals how historical urban layers remain legible within the morphology of the contemporary city, demonstrating the continuous reinterpretation and adaptation of former defensive infrastructures over time.

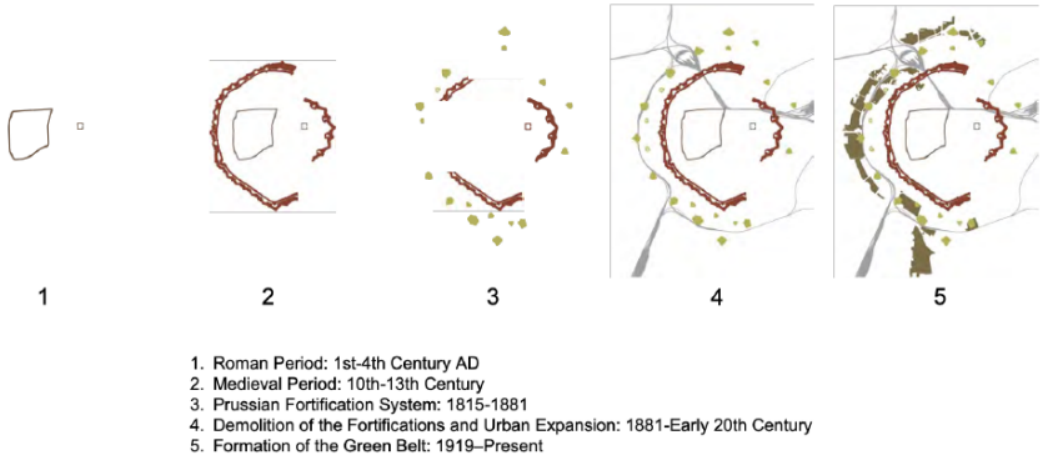


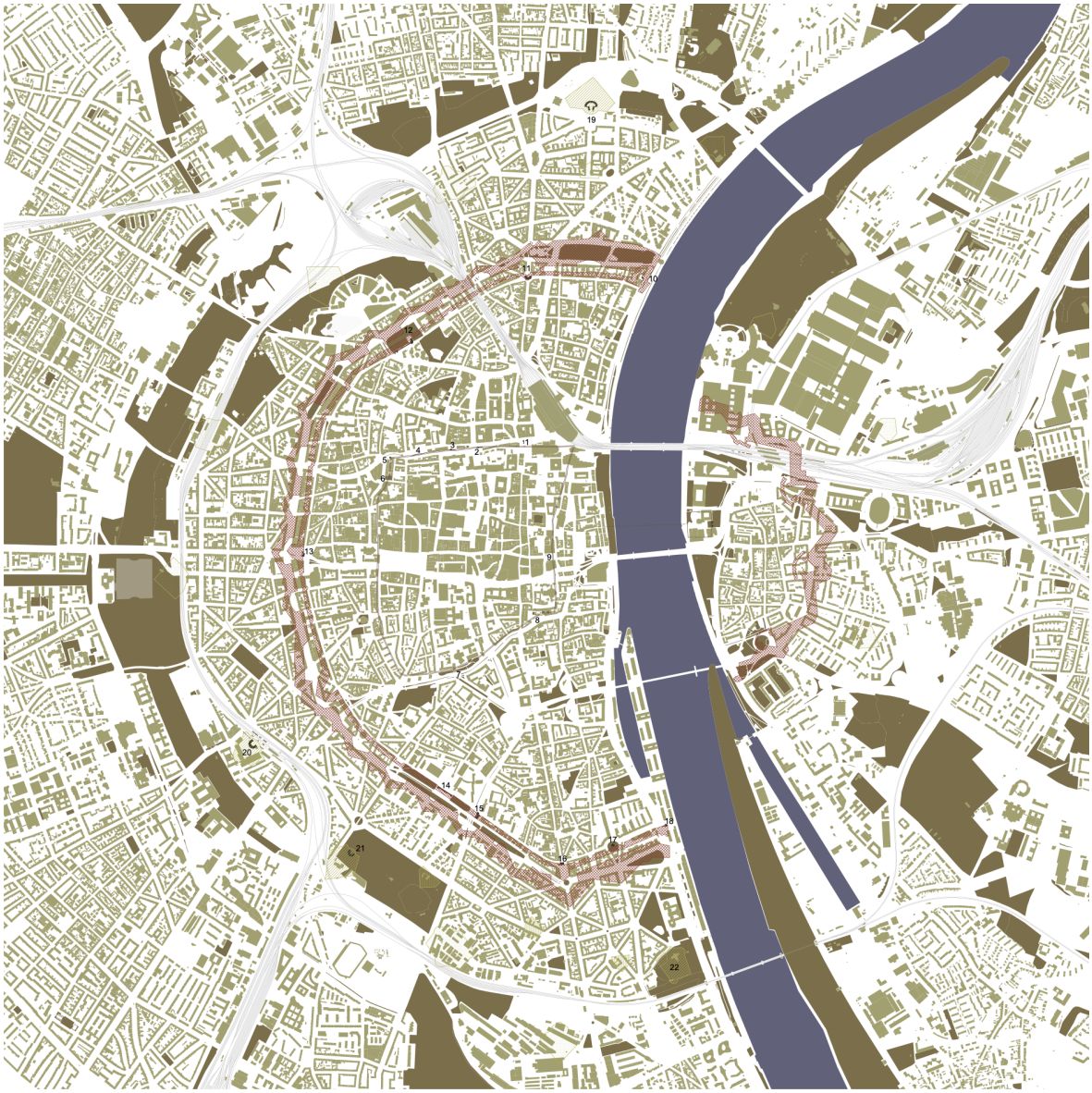
Fig. 1. Historical evolution of Cologne's urban fortification systems and the formation of the Green Belt, from the Roman city to the contemporary urban structure

3. Historical Evolution of Cologne's Urban Structure: From the Roman era through the Middle Ages to the present day

To better understand the historical formation of Cologne's contemporary urban structure, Figure 2 overlays the city's historical defensive systems onto the present-day urban map. Rather than reconstructing Cologne at different moments in time, the drawing reveals how successive historical layers remain embedded within the contemporary city. By superimposing the Roman city wall, the medieval fortifications, the Prussian fortress system, surviving architectural fragments, and the Green Belt onto the current urban fabric, the map illustrates the spatial continuity of Cologne's historical development. It also demonstrates how historical boundaries, despite the loss of their original functions, continue to influence the organisation and morphology of the city.

The Roman city wall originally defined the boundary of *Colonia Claudia Ara Agrippinensium* and established the first spatial framework of Cologne. Although only isolated fragments survive today, its former alignment remains legible in parts of the street network and urban blocks, representing the city's oldest spatial layer. During the Middle Ages, rapid urban growth required the construction of a much larger fortification system, which enclosed the expanding city while incorporating the former Roman settlement. Following the demolition of most medieval walls after 1881, their course was transformed into today's Ring roads, which continue to define the transition between the historic centre and the nineteenth-century urban expansion.

In the nineteenth century, the Prussian military established a new defensive ring of detached forts beyond the medieval city. These forts and their surrounding military zones created another urban boundary while simultaneously preserving extensive open spaces around Cologne through strict building restrictions. Following the First World War, many former fortification sites and military open spaces were converted into public parks and green corridors, forming the foundation of Cologne's Green Belt. Several forts, gates, towers, and wall fragments remain preserved today and continue to function as historical landmarks within the contemporary city. Figure 2 demonstrates that Cologne's present urban structure is the result of successive historical transformations, in which each urban layer has been continuously adapted and remains legible within the contemporary urban fabric.



Former Roman City Walls
Former Roman Bridge
Former Roman Forts
Former Medieval City Walls
Former Fortifications, Partially Preserved and Still in Use
Former Fortifications, Partially Preserved, No Longer in Use
Former Fortifications, Preserved Only as Urban Traces
Former Fortifications, Demolished
Built-up Area
Green Space
Rhine River
Fragments of Historic Urban Elements

Fragments of the Roman City Walls

1. Roman North Gate
2. Roman City Wall
3. Roman Fountain
4. Roman Tower
5. Römerturm
6. Helen's Tower
7. Roman City Wall
8. Roman Wall
9. Roman City Wall

Fragments of the Medieval City Walls

10. Kunibert Tower
11. Eigelstein Gate
12. Medieval City Wall
13. Hahnen Gate
14. Prinzen-Garde Tower & Saxon Tower
15. Ulreforte
16. Severin Gate
17. Bott Mill
18. Bayen Tower

Fragments of the Fortifications

19. Fort X - Wilhelm
20. Fort V
21. Fort IV - Paul
22. Fort I - Rheinschanze

Fig. 2. Overlay of Historical Urban Fortifications and Their Surviving Fragments within Contemporary Cologne

4. Conclusions

This study has demonstrated that Cologne's contemporary urban structure is the result of successive historical transformations rather than a single planning intervention. The Roman city wall, the medieval fortifications, the Prussian fortress system, and the Green Belt should not be understood as isolated historical elements but as interconnected layers that continue to shape the city's morphology. Although many defensive structures have disappeared physically, their spatial logic remains legible through streets, public spaces, green corridors, and surviving architectural fragments. The cartographic analysis presented in this study reveals that historical urban boundaries have continuously been reinterpreted and integrated into the evolving city. Rather than representing the end of former defensive systems, the demolition of fortifications created new opportunities for urban expansion, transportation infrastructure, and the development of the Green Belt. Today, the Green Belt remains one of Cologne's most important urban structures, despite its fragmented condition. Future planning and regeneration projects should therefore be based not only on functional or environmental considerations but also on a comprehensive understanding of the city's historical spatial structure. Reading the city through its historical layers provides a valuable framework for reconnecting fragmented urban spaces while strengthening Cologne's spatial identity and cultural continuity.

References

1. Meyen, H.: Die Kölner Grünanlage. Schwann, Düsseldorf (1979).
2. Meyen, H.: Die Kölner Stadtbefestigungen. Einzigartige Zeugnisse aus Römerzeit, Mittelalter und Neuzeit. Regionalia, Daun (2021).
3. Meyen, H.: Eine Grünanlage mit Geschichte. Festungsbauten und Äußerer Grüngürtel in Köln. Maternus, Köln (2015).
4. Bauwens-Adenauer, P., S. Soénius, U.: Der Masterplan für Köln: Abert Speers Vision für die Innenstadt von Köln. Greven, Köln (2009).
5. Fritz, S., Wilhelm, A.: Köln – Entwicklungsfragen einer Großstadt. Saaleck, München (1923).
6. Jansen, H., Ritter, G., Wiktorin, D., Weiss, G., Gohrbandt, E.: Der historische Atlas Köln: 2000 Jahre Stadtgeschichte in Karten und Bildern. Emons, Köln (2003).

Value Analysis in the World Heritage Convention : recognition and interpretation of the 'Criterion (ii)' on interchange of human value.

Semina An[1]

[1] IWTG, Silbermannstraße 2, 09599 Freiberg, Germany
Ssam2na@gmail.com

Abstract. Criterion (ii) of the 10 World Heritage Criteria describes heritage as the result of "interchange" between different cultures. The definition of criterion (ii) was modified from "influence" to " interchange" by the 1996 Operational Guidelines. It is understandable that the view of World Heritage on cultural heritage has also changed from "one direction" cultural impact to "cultural interaction". Half a century after the adoption of the World Heritage Convention, it is necessary to review whether this criterion is properly applied to cultural heritage.

But, for a considerable number of listed sites, there is a gap in the recognition and interpretation of the value of criterion (ii) and the site. This paper recognizes the basic notion of "exchange of human values", which has not been studied before and uses the evolutionary chronology of criterion (ii) to interpret the intent of the change. The interpretation identifies the problem and discusses future considerations for criterion (ii).

Keywords: The World Heritage Convention, Criterion (ii), Influence, Interchange of Human Values.

1. Introduction

This year marks the half-century since the adoption of the World Heritage Convention in 1972, and it is time to review the traces of the past and prepare for the next 50 years. In particular, it is necessary to verify that the World Heritage criteria, which are a standard grid for describing the OUV of a property, have been properly stated on the site.[2, 3]

The criteria of the property are specified by the State Party in the nomination dossier and evaluated by the World Heritage Advisory Bodies, ICOMOS or IUCN, and the World Heritage Committee adopts the final decision on the OUV of the property based on the recommendations of the Advisory Bodies. Understanding and defining the criteria for the listing is critical, as it not only has a significant impact on the nomination but is also directly related to the direction of the heritage conservation and management plan.

Another important reason for the study of the criteria is the fact that the definition of the criterion has been modified on the basis of the experience accumulated in the last 50 years, for example, in living beings. Thus, due to insufficient study and understanding of the revised criteria, errors have often been made in explaining the inherent value of heritage through the criteria.

Among the ten criteria, especially criterion (ii) has been changed its main notion radically. The present definition of criterion (ii) has been modified a total of five times in the operational guidelines.[4] The significant difference in the definition is that "influence" was changed to "interchange" in 1996 when it was last revised for criterion (ii). The "interchange of human values" in criterion(ii) is the only reciprocal value resulting from the interactions and multidimensional dialogue between countries, regions or cultural clusters. In other words, heritage that meets criterion (ii) must be considered in the transnational dimension, not only in the field but in integrating relations with other cultures and countries relating the interchange value in terms of sustainable development and resilience.

[2]Para. 5 of Article 11 in the World Heritage Convention states that 'the Committee shall define the criteria on the basis of which a property belonging to the cultural or natural heritage may be included in either of the lists mentioned in paragraphs 2 and 3 of this article.'

[3]Decision CONF 001 VI.A(a).18 (whc.unesco.org/en/decisions/2038/, Accessed to 26 August 2022)

[4]UNESCO WORLD HERITAGE CENTER, The Operational Guidelines for the Implementation of the World Heritage Convention, 1977 ~2019.

2. Evolution of Criterion (ii)

2-1. ICOMOS report on World Heritage Criteria (1976)

What was the reason and trigger for changing the description of criterion (ii)? The starting point for the establishment of criterion (ii) is the "UNESCO informal consultation of intergovernmental and non-governmental organizations in the implementation of the Convention" in Morges in May 1976.5 At this meeting, ICOMOS proposed the first draft of criteria (i) ~ (vi), and criterion (ii) was stressed "influence over the development". The report also provides some examples of potential properties relating each criterion to understand the criteria. As indicated by ICOMOS' descriptions and examples, the significance of criterion (ii) must be a heritage that impact other cultures.

2-2. Changes to Operational Guidelines until 1994

Since the first adoption of criterion (ii), which reflected the ICOMOS proposal, the basic idea of the criterion as "influence" remained until the 1994 Operational Guidelines.[6]Criteria, as with the evolution of living organisms, have been revised several times by discussions among the Committee members, experts, and advisory bodies.

Compared to the ICOMOS report, the sentence structure of criterion (ii) in the Operational Guidelines (1977) was largely changed, and the objects of sites are refined in the guidelines and were added "monumental sculpture, garden and landscape design, related arts". According to the final report of the 1st session of the Committee, it was decided to delete examples that could affect the criterion to evaluate cultural properties as they could cause prejudice on the site.[7] Through this process, it can understand the attempts of the Committee to maintain equality and fairness in the evaluation of sites.

In the 1978 Operational Guidelines, the term "subsequent" was omitted and kept consistently concise in this sense. The term "town-planning" was added to the objects that can be assessed under criterion (ii). [8]

However, the definition of criteria remained difficult, as the wording of the Convention itself was occasionally imprecise. Therefore, the Bureau of the Committee asked Mr. M. Parent to confirm a detailed criteria study to define a property.[9] One of in his paper, he pointed out that criteria (ii), (iv) and (v) needed the comparative assessment,[10] and the results of his study were then reflected in a revised version of the 1980 Operational Guidelines.[11] The definition of criterion (ii) in the Guidelines was slightly modified. Compared to the 1979 definition, the wording "considerable influence" was changed to "great influence". The category of the property to be evaluated was modified to "architecture, monumental arts, or town-planning and

---

[5]The UNESCO informal consultation held with ICCROM, ICOMOS, IUCN, and UNESCO Secretariat on the implementation of the Convention. They discussed three aspects: Criteria for the inclusion of property in the World Heritage List; Format and content of documentation to be requested from States; and Criteria for the determination of an order of priorities for awarding international assistance. ICOMOS, *Proposals made by the International Council on Monuments and Sites, in UNESCO, Informal consultation of intergovernmental and non-governmental organizations in the implementation of the Convention concerning the protection of the world cultural and natural heritage*, Morges, 19-20 May 1976. (CC-76/WS/25. annex III)
UNESCO WORLD HERITAGE CENTER, *Final report on the informal consultation of intergovernmental and non-governmental organisations on the implementation of the convention concerning the protection of world cultural and natural heritage*, 1976.

[6] JOKILEHTO JUKKA, *The World Heritage List. What is OUV? Defining the Outstanding Universal Value of Cultural World Heritage Properties*, Berlin, hendrik Bäßler, 2008.

[7]UNESCO, *final report*, In "Intergovernmental Committee for the protection of the world cultural and natural heritage". Paris. 17 October 1977, Para. 29. (CC-77/CONF.001/9)

[8]The 'Medina of Fez' in Morocco was inscribed on the List in 1981 with criteria (ii) and (v). The heritage is described its criterion (ii) as being influenced by the city of eastern Mediterranean on town-planning development. It was the first inscription by recognizing the value of town-planning has been written in the criterion. WHL, Medina of Fez (https://whc.unesco.org/en/list/170, Accessed 09 July 2021)

[9]UNESCO, *Convention Concerning the Protection of the World Cultural and Natural Heritage. Third Session of the World Heritage Committee*. Paris. October 1979. Para. 3 (CC-79/CONF.003/11)

[10]MICHEL PARENT, *Comparative Study of Nomination and Criteria for World Cultural Heritage. Paris. September 1979.* (CC-79/CONF.003/11 ANNEX)

[11]UNESCO, *Operational Guidelines*. Paris. October 1980. WHC/2 Revised. Para. 19.

landscaping". The phrase "monumental sculpture … related arts" was merged into "monumental arts", broadening the meaning of the spectrum. The phrase "garden and landscape design" also merged into "landscaping" but was changed back to "landscape design" in the Guidelines (1994). The defined description was maintained until 1994.

2-3. World Heritage Expert Meeting on Global Strategy, Paris (1994)

The year 1994 can be recognized as a turning point about criterion (ii), as the framework for World Heritage was changed by the Expert Meeting on Global Strategy, the Nara Documents, and the Expert Meeting on Heritage Canals in Canada. The Committee launched the Global Strategy for a representative, balanced and credible World Heritage List. Its aim is to ensure that the List reflects the world's cultural and natural diversity of OUV.[12]

The work group of the Global Strategy recommended the revision of criteria for cultural heritage.[13] About Criterion (ii) had comment as "Re-examine this criterion so as to reflect better the interaction of cultures, instead of the present formulation, which suggests that cultural influences occur in one direction only". In the regards of this comment, "the Delegate of Thailand stressed the importance of the proposed modification to criterion (ii) which, in its present form, only takes into account cultural influences which occurred in one direction only, between the different continents. This re-examination should be carried out in-depth in order to avoid all notion of any cultural domination: efforts must be made to achieve diversity of the manifestations of different cultures in their interaction with their environment".[14] Right after his comment, the Delegate of China mentioned as "to continue according to attention in the future to the diversity of cultures so as to achieve a fully representative List". At the 18th session of the Committee, the criterion was finally redefined as it is today, referring to the work group's recommendations for modification.[15]

3. Analysis of the final definition of criterion (ii): an important interchange of human values

The definition of criterion (ii) in the Operational Guidelines in 1996 was reflected by the comprehensive results of the previous expert meetings and the Committee. To analysis the justification of the criterion, first, the Cambridge dictionary definition of "interchange" states that it is an exchange between different people, especially intangible things, ideas, and information. As the revision from "Influence" to "Interchange", the flow of culture is also multiple direction. Human values generally refer to the values that are at the core of being human and are considered fundamental, inherent values of human beings, including truth, honesty, loyalty, love, peace, and so on. This term is usually used in the Constitution or professional ethics. However, the term used for criterion (ii) stands for "ideas" according to the presentation material produced by Ms. Susan Denyer at the international world heritage expert meeting on criterion (vi) in Warsaw, Poland in 2012. But the meaning of human values needs to identify more. Lastly, the objective category was expanded from only "architecture or human settlements" to "monumental art or landscape design", after the discussion of the historic canal in Canada (1994), the term "technology" is added and recognized as a value of interchange. The spectrum of value was broader than before.
Currently, the definition of criterion (ii), as amended in 1996, is retained as the latest version.

[12]Official definition of "Global Strategy" on the website of World Heritage Centre (whc.unesco.org/en/globalstrategy, Accessed 12.08.26)
[13]UNESCO, *Expert Meeting on the "Global Strategy" and thematic studies for a representative World Heritage List* (UNESCO Headquarters, 20-22 June 1994), In "the 18th session of Bureau of the World Heritage Committee". UNESCO Headquarters, Paris. 28 June 1994. WHC-94/CONF.001/INF.4
[14]"XII. Global Strategy for a Representative World Heritage List."
UNESCO: *Report of the Rapporteur*. In "18th session of the Bureau of the World Heritage Committee" (UNESCO Headquarters, Paris, 4-9 July 1994). 19 August 1994. (WHC-94/CONF.001/10, 19)
[15]UNESCO, *Convention Concerning the Protection of the World Cultural and Natural Heritage, World Heritage Committee*, Eighteenth session, Phuket, Thailand, 12-17 December 1994, Report (WHC-94/CONF.003/16, 31 January 1995)

4. Problems the recognition and interpretation of criterion (ii)

It has always been a challenge to determine the value of heritage at different sites. As a result, there are potential errors in the interpretation of the State Party's inscription criteria, and the value of the cultural heritage may not be properly recognized. In fact, some sites that were in transition to changing the definition of criterion (ii) provided confusing explanations about the meaning of the criterion.

For instance, Historic Fortified City of Carcassonne was inscribed on the list with criteria (ii) and (iv) in 1997. The description of this site's criterion (ii) "... It is of exceptional importance by virtue of the restoration work carried out in the second half of the 19th century by Viollet-le-Duc, which had a profound influence on subsequent developments in conservation principles and practice." Although the main notion of criterion (ii) at the time was an interchange value, it is still described as an influence. If its inherent value is not properly understood, this can also lead to mistakes in maintaining the value of the property in the future.

5. Conclusion and plan of this research

The evolution of the criteria over the 50-year history of the Convention is the result of changing perceptions of World Heritage over that time. This year is the appropriate time that we look to the last half century for the future of World Heritage.

There is a saying that the process and the results are good only when the first button is well tightened. If the expression of the OUV of the property are inconsistent with the criteria, there will be a misalignment in the overall process of the World Heritage Site between the application, the evaluation, the listing and the conservation management, which will seriously affect the sustainability and resilience of the World Heritage Sites.

Criterion (ii) has not been studied till now, so in many cases, the "interchange of human values" has been confused with the concept of "influence" before the amendment. This means that when we need to discuss sustainable development and resilience, an unstable cornerstone has been laid. At this point, we need to examine the criteria.[16]

The development of this study will analyze criterion (ii) based on case studies that have been inscribed as World Heritage. It may also be an interesting experiment to see if the criteria for inscription can be re-evaluated with the revised criterion (ii), "interchange of human values", among properties inscribed before the 1996 revision. The best examples of this possibility study are the Galle Fortress in Sri Lanka, inscribed with criterion (iv) in 1988, and the Imperial City of Hué in Vietnam, inscribed with criterion (iv) in 1993. In addition, ICOMOS annual reports on "World Heritage Evaluations" are used in this research, especially when the ICOMOS recommendation is "not to inscribe" or "deferred" and the file was subsequently withdrawn by the States Parties. The analysis of these cases, whose evaluations are not published on the World Heritage Centre website, helps for the depth study of criterion (ii) in the form of a multi-layered analysis.

This work is important not only for the study of World Heritage criterion (ii) but also to complete the identification of World Heritage and to ensure a truly representative World Heritage List and thus its credibility for the next 50 years. And also, I would like to ask the question for next step: "do we need to complete criterion (ii) today to be consistent with its actual use?"

[16] Criterion (vi) had been studied: JUDITH HERMANN, *Tracing change in World Cultural Heritage: the recognition of intangible heritage.* Ph.D. Thesis, Faculté de l'aménagement, Université de Montréal, 2015. Available at: papyrus.bib.umontreal.ca/xmlui/handle/1866/14112 CHRISTINA CAMERON; JUDITH HERMANN, *Guidance and capacity building for the recognition of associative values using world heritage criterion (vi): Final Report.* World Heritage Centre Paris, 2018 (Document-327-20)

References

1. UNESCO WORLD HERITAGE CENTER, Final report on the informal consultation of intergovernmental and non-governmental organisations on the implementation of the convention concerning the protection of world cultural and natural heritage, 1976.
2. UNESCO WORLD HERITAGE CENTER, The Operational Guidelines for the Implementation of the World Heritage Convention, 1977 ~2019.
3. JOKILEHTO JUKKA, The World Heritage List. What is OUV? Defining the Outstanding Universal Value of Cultural World Heritage Properties, Berlin, hendrik Bäßler, 2008.
4. CHRISTINA CAMERON; MECHTILD RÖSSLER, Many Voices, One Vision: The Early Years of the World Heritage Convention. Farnham, 2013.
5. JUDITH HERMANN, Tracing change in World Cultural Heritage: the recognition of intangible heritage. Ph.D. Thesis, Faculté de l'aménagement, Université de Montréal, 2015. Available at: https://papyrus.bib.umontreal.ca/xmlui/handle/1866/14112

MLK Models and Industrialized Architecture in the GDR

Sung-Yong Kim[1]

[1] Kustodie, IWTG, TU Bergakademie Freiberg, Lessing Straße 45, 09599 Freiberg, Germany
yongcocco@gmail.com

Abstract. This paper explores the phenomenon of "Industrialized Architecture in the GDR" (IndustrialisierteArchitektur in der DDR) by examining the architectural models and building systems of the *Metalleichtbaukombinat Leipzig* (MLK), an East German state-owned enterprise active from 1969 to 1990. While existing scholarship on East German architecture has predominantly focused on residential mass housing such as the *Plattenbau*, this study shifts the analytical lens toward commercial and industrial architecture, utilizing a newly rediscovered collection of approximately 80 MLK architectural models housed at the TU Bergakademie Freiberg.

Set against the backdrop of the GDR's postwar economic constraints and its centralized planned economy (*Planwirtschaft*), the study investigates how the state pursued advanced architectural industrialization through prefabrication, typologization, and standardization to maximize construction efficiency and generate vital foreign exchange via architectural exports (*Architekturexport*). By analyzing the historical trajectory and core characteristics of the "industrialization of building" (*Industrialisierung des Bauens*), this research illuminates how MLK's lightweight steel-frame structures materialized state ideology and global trade ambitions simultaneously. Finally, the paper addresses the complex heritage dilemmas surrounding these aging structures, evaluating the criteria for their preservation value and discussing how the architectural legacy of the GDR offers critical precedents for future preservation discourse in divided nations.

Keywords: Industrialization of Architecture, GDR Architecture & MLK Models, Preservation Criteria of Industrialized Architecture

1. Introduction

Even if you get lost in Berlin, it's easy to find your way again. While the wide, well-organized streets certainly play a part, it's also because the Berlin TV Tower (Fernsehturm)—erected in the heart of the city, much like Seoul's Namsan Tower—stands like a lighthouse in the distance, guiding the way. Located in Park am Fernsehturm in Berlin-Mitte, this massive transmission tower stands 368 meters tall and is still recognized as the tallest structure in Germany. However, many visitors—including Berliners themselves—are unaware that this structure was built during the era of the German Democratic Republic (Deutsche Demokratische Republik, 1949–1990; hereinafter DDR). Furthermore, it is even less widely known that the structure's highlight—the circular observation deck—was built by the East German industrial combine Metalleichtbaukombinat Leipzig (hereinafter MLK). Alongside such iconic structures, countless standardized buildings were constructed throughout the former East Germany, and most of these buildings are now in need of maintenance and repair or are at risk of demolition. However, under the political label of "socialist architecture," there has not been sufficient discussion of their architectural characteristics.

While residing in Freiberg, a small city in the state of Saxony that was formerly part of East Germany, I was introduced to the fields of industrial archaeology and industrial culture (*Industriearchäologie / Industriekultur*) at the TU Bergakademie Freiberg and had the opportunity to participate in a project analyzing architectural models created during the GDR era. These models were produced between 1970 and 1990 at the MLK model workshop for various domestic and international exhibitions and fairs; they are characterized by being classified into types based on their diverse structures, materials, and purposes. (Figure 1) To address chronic financial difficulties and increase its foreign exchange reserves, the East German government exported various goods to several countries, including those in the Eastern Bloc; one of the most representative export items was architectural exports (*Architekturexport*)[1], and MLK was a key enterprise in the export of steel-frame buildings.

It was at this point that one of my preconceptions about architecture was shaken. The idea that architecture could be exported like any other product was a fact that had never crossed my mind. However, East German architecture was indeed typologized, standardized, and exported throughout East Germany and abroad. Since this was a concept rarely encountered in traditional architectural history, I initially struggled to find precise

words or concepts to describe it. The notions of architectural typology, mass production, and standardization were subsequently recognized as components of the broader concept of “the industrialization of building”

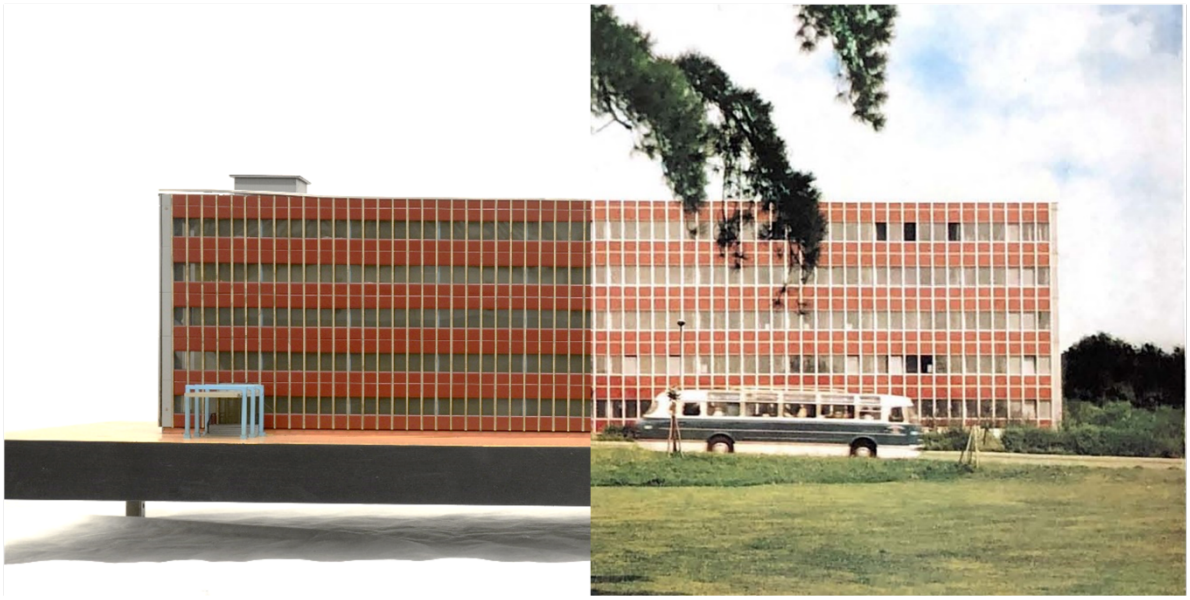


Fig. 1. The Model and the realized building of "Type Leipzig" (©Kustodie, TU Bergakademie Freiberg)

(*Industrialisierung des Bauens*) [2], which ultimately led me to the overarching topic of “Industrialized Architecture in the GDR” (*IndustrialisierteArchitektur in der DDR*). So, how can industrialized buildings—which are typologized and mass-produced—be differentiated from traditional approaches to architectural preservation? If the concept of the “industrialization of architecture” can effectively explain the built environment of the GDR, what role might it play in current discussions on the preservation of GDR architecture?

2. Historical Background

To understand architecture in the GDR, one must first examine the political and economic circumstances facing the state at the time. As is well known, Germany emerged as a defeated nation from World War II (September 1, 1939–September 2, 1945), and with its major cities and industrial facilities severely devastated, the country found itself in a state of national crisis. In East Germany in particular, the damage was substantial: approximately half of the residential buildings in some cities and about 40% of industrial facilities were destroyed. Consequently, the postwar housing shortage and a systematic response to it became critical and urgent priorities for both German states. In West Germany, stable reconstruction was facilitated through full-scale financial support from the United States and the procurement of materials from the capitalist bloc. In contrast, during the early stages of reconstruction in East Germany, a significant portion of its production facilities had to be dismantled and handed over to the Soviet Union as war reparations, with most of these obligations settled only by 1960. Even afterward, the economy suffered from chronic financial difficulties due to restricted trade with the Western bloc, technological stagnation, and declining productivity.[3]

Concurrently, both East and West Germany faced intense pressure to rapidly rebuild urban infrastructure and provide housing. To overcome this challenge, East Germany focused on standardizing and dividing labor in construction to maximize efficiency. First, architectural elements, functions, structures, and technical requirements were synthesized and categorized, driving the industrialization of architecture at the national level. Since the early 20th century, as the global pursuit of inexpensive, rapid—so-called “efficient”—construction gained momentum, the industrialization of architecture was steadily attempted not only in socialist countries, including East Germany, but also in various capitalist nations.[4] However, in the case of East Germany, anchored in its socialist political-economic system, national resources were concentrated

heavily on the typification and standardization of architecture, resulting in an exceptionally high degree of architectural industrialization. As expressed by Nikita Khrushchev, then First Secretary of the Communist Party of the Soviet Union, this state-led approach is best encapsulated by his directive: "Let's build better, cheaper, and faster" (*"Besser, billiger und schnellerzubauen."*).[5]

Significantly, the "planned economy" (*Planwirtschaft*) system in East Germany—heavily influenced by the Soviet model—made state control over political and economic spheres both uniform and top-down, while ensuring systematic organization and efficiency. All enterprises operated under the direct oversight of the Socialist Unity Party of Germany (*SozialistischeEinheitsparteiDeutschlands*, hereinafter SED), transforming cities and buildings from private assets into state and party property. Beginning in the 1960s, virtually all economic entities in East Germany became state-owned enterprises (*Volkseigene Betriebe*, hereinafter VEB) and their combined bodies known as *Kombinate*. These served as the fundamental corporate vehicles of the planned economy, acting as the institutional instruments that brought party ideology into physical reality.[6]

Under this centralized planned economy, the design and construction of urban and architectural projects were intensively "industrialized" by the state. This socialist approach to construction generated a uniform, community-centered urban landscape across East Germany. In particular, the industrialized residential architecture of the period—represented by the *Plattenbau* (large-panel system building)—remains a core element of the East German urban landscape. As architectural symbols of socialism, these structures are easily identifiable not only in the former East Germany but throughout the broader Eastern Bloc.

In addition to residential buildings, the East German government standardized commercial and industrial structures; among these, lightweight steel-frame buildings were pioneered by MLK. MLK operated as the state-owned enterprise responsible for steel-frame construction in East Germany from 1969 to 1990. Headquartered in Leipzig, it served as a key national consortium dedicated to large-scale steel and metal construction. It was responsible for the structural steelwork of major landmarks such as the Berlin TV Tower, the *Palast der Republik*, and the *Gewandhaus* in Leipzig, as well as crucial industrial infrastructure—including power plants and large bridges—and medium-sized commercial facilities like greenhouses, schools, and offices. Specializing in high-efficiency and technologically sophisticated steel-frame construction, MLK stood as a leading player in the non-residential construction sector.[7]

Politically, the East German government used architecture to give tangible form to socialist ideology. If cities and buildings are understood as materializations of the sociocultural contexts of their creation [8], the urban landscapes of East Germany and the Eastern Bloc can be regarded as the physical embodiments of socialist ideals. In other words, buildings distributed to the public equally and fairly in accordance with socialist principles functioned simultaneously as visual propaganda for state ideology. Under this political environment, structures across all industrial sectors—including residential architecture—were systematically standardized and industrialized, reflecting both the demands of the era and the state's instrumentalization of those demands.

3. MLK Architectural Models and the Industrialization of Architecture

These models, left untouched for over two decades deep within the storage of the Kustodie—the central university collections department—at the TU Bergakademie Freiberg, returned to public view in 2020. Distinguished by their vivid colors, high level of craftsmanship, and numbering approximately 80 models of varying sizes, these artifacts vividly capture the architectural characteristics of the East German era. The following year, in 2021, a research project titled *Reconstruction and Analysis of the Historical Development of Steel and Light Metal Construction in the GDR Using the Model Collection of the MLK Leipzig* was selected and funded by the German Research Foundation (*Deutsche Forschungsgemeinschaft*, DFG), running from 2021 to 2023.[9]

Consequently, the actual structures erected by MLK began to be re-examined as the materialized counterparts of these models. Beyond the models and their corresponding building typologies, the overarching history of MLK—particularly its extensive international activities—was brought back into sharp focus.

The MLK models serve as crucial historical artifacts that encapsulate the preceding discussions. They reflect the postwar public demand for efficient and rapid urban reconstruction; the realization and propagation of state socialist ideology; the simplification, standardization, and industrialization of architecture driven by

severe financial constraints; and the phenomenon of architectural export. The MLK models lie precisely at the intersection of all these historical dimensions. (Figure 2)

Furthermore, as scholarly discourse expands from the industrialization of residential architecture—historically the primary focus—to encompass industrial and commercial facilities, MLK's models hold significant value as compelling empirical evidence.

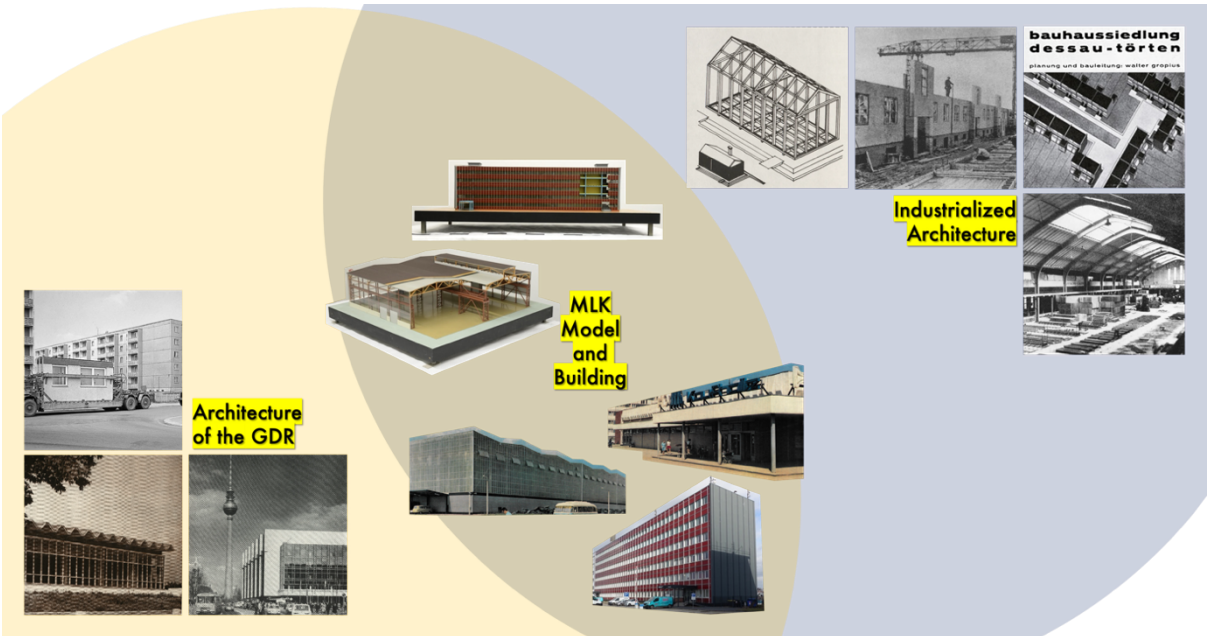


Fig. 2. The MLK Model and the Building within the Context of the Relationship Between Industrialized Architecture and GDR Architecture (©Sung-Yong Kim)

4. Discussion on the Preservation Value of Industrialized Architecture, Including MLK Buildings

Despite the characteristics outlined above, East Germany's standardized architecture leaves room for ongoing debate regarding its preservation value. The crux of these discussions centers on the argument that while sharing broader methodological contexts with traditional architectural heritage conservation, distinct preservation criteria must be established specifically for industrialized architecture, which was mass-produced using uniform designs.

The industrialized architecture of East Germany is particularly vulnerable under traditional doctrines of architectural uniqueness. Among identical or highly similar structures, an "original" building often does not exist or is difficult to define. While one might intuitively expect heritage value to increase as surviving examples decrease, preservation value in this context cannot be strictly inversely proportional to the remaining count. Many GDR buildings were erected en masse according to predefined types using standardized components. Consequently, preserving every building of a given type is practically impossible, and selecting which specific examples warrant preservation remains a complex challenge.

Nearly 40 years after the German reunification in 1989, buildings from the East German era are aging, prompting active discussions regarding their cultural heritage value. At the center of this debate is the *Plattenbau*, the quintessential typified residential architecture of East Germany. Although preservation efforts encounter challenges due to constrained construction quality, the significance of these buildings as living spaces and repositories of memory for former GDR citizens continues to be recognized. Similarly, MLK commercial and industrial buildings face the reality of unexamined demolition, driven by material deterioration and a general lack of public recognition. However, these structures—which are indispensable to understanding the totality of East German architecture—must be integrated into public discourse promptly, accompanied by clear and rigorous criteria for assessing their preservation value.

Rather than viewing GDR architecture merely as the somber legacy of a failed political regime, efforts should be made to understand it within the broader framework of 19th- and 20th-century architectural industrialization, while fully accounting for the diverse characteristics born from the unique political conditions of the GDR. Moreover, given South Korea's unique position as a divided nation, discussions surrounding the preservation of GDR architecture will serve as a vital precedent for future approaches to the preservation of North Korean architecture following potential reunification—a task that will demand immense sensitivity and historical care.

References

1. Butter, A., Flierl, T. (eds.): Architekturexport der DDR. Zwischen Sansibar und Halensee. Lukas Verlag, Berlin (2023).
2. Schmid, T., Testa, C.: Bauen mit Systemen. Verlag für Architektur Artemis, Zurich (1969).
3. Sieber, F., Fritsche, H.: Bauen in der DDR. Huss-Verlag, Berlin (2006).
4. Nerdinger, W.: Architektur in Deutschland im 20. Jahrhundert. Prestel, Munich (2023).
5. Flierl, T.: 'Possibly the greatest task an architect ever faced—Ernst May in the Soviet Union (1930–1933)'. In: Quiring, C., Voigt, W., Schmal, P.C., Herrel, E. (eds.) Ernst May 1886–1970, p. 157. Prestel, Munich, London, New York (2011).
6. Federal Agency for Civic Education (bpb): Planwirtschaft,

   https://www.bpb.de/kurz-knapp/lexika/politiklexikon/18008/planwirtschaft/, last accessed 2026/08/05.
7. Tasche, M., Baumgarten, H., Zinnert, G.: Standardized Steel Beam-Column Structures in the GDR from 1955 to 1990. Bautechnik **88**(8), 521 (2011).
8. Hoscislawski, T.: Building Between Power and Powerlessness. Berlin (1991).
9. DFG Project homepage: Stahl- und Metallleichtbau in der DDR, https://kulturerbe-konstruktion.de/spp-2255-teilprojekt/stahl-und-metallleichtbau-in-der-ddr-c4/, last accessed 2026/07/30.

The 15-Minute Play City: Infrastructure for Friction and Resonance in the Age of AI

Jaewook Kwon

Earthscape Inc., Playground Designer, Brandenburg an der Havel, Germany
shius2@gmail.com

Abstract. Debates on children, smartphones, and generative artificial intelligence (AI) often centre on age limits, phone-free schools, and limits on screen time. Such restrictions may reduce digital exposure but cannot restore the embodied, social, and spatial experiences displaced by a phone-based childhood. This paper proposes the 15-Minute Play City as public infrastructure for such experiences. Developmental friction denotes manageable physical, social, cognitive, and temporal resistance through which children practise agency; civic friction denotes repeated encounter and shared work through which residents negotiate difference and responsibility. Hartmut Rosa's concept of resonance provides a normative criterion: people can be affected, respond in their own voice, and be changed without controlling the encounter. The model's cross-cutting principle, experience before assistance, does not prescribe a sequence for using AI during play. It is an urban priority: children need bodily exploration and reciprocal human encounter before AI becomes a default provider of companionship, interpretation, or experience. The present concern is digital displacement; future ambient AI may also mediate presence itself.

Keywords: 15-Minute Play City, developmental friction, civic friction, resonance, child agency, artificial intelligence.

1. Introduction: From Digital Control to Urban Provision

Jonathan Haidt describes a shift from a play-based to a phone-based childhood, alongside declines in independent mobility, unsupervised peer interaction, sleep, and sustained attention [1]. Causal claims require caution: associations between social media and youth mental health vary by age, vulnerability, content, and context [2]. Rather than claiming smartphones alone cause a mental-health epidemic, this paper asks which developmental experiences are crowded out when digital environments become the easiest places to act, socialise, and escape boredom.

Anna Lembke's account of immediate reinforcement helps explain why systems that minimise waiting, frustration, and disagreement can become attractive substitutes for displaced experience [3]. Generative AI may intensify displacement through personalised content and synthetic companionship with little risk of rejection or obligation. This is a design concern, not proof that AI causes harm. Noreena Hertz frames loneliness as feeling unseen, unheard, and detached from citizens and institutions [4]. Their roles are distinct: Haidt diagnoses displaced experience; Lembke explains frictionless reward; Hertz identifies the relationships to rebuild.

First, the paper reframes digital childhood from a matter of private screen management to a problem of unequal public infrastructure. Second, it links developmental and civic friction with Rosa's resonance without romanticising risk or conflict. Third, it extends the 15-minute city from adult-measured proximity to child-usable access, and the playground from equipment to life-course social infrastructure. Experience before assistance becomes a two-horizon principle: resist digital substitution now and protect unmediated encounter as ambient AI develops.

2. From Productive Friction to Resonance

Developmental friction is manageable, legible, and recoverable resistance through which children practise regulating their bodies, attention, emotions, judgement, relationships, and time. Height, balance, loose materials, peer conflict, route-finding, and unstructured time can support multiple forms of agency when children can adjust difficulty, anticipate consequences, retry, and obtain help. It excludes neglect, humiliation, preventable hazard, and inaccessible design. Civic friction is the productive encounter with difference through

shared tasks, co-making, care, and negotiation. Co-presence alone is insufficient; equal status, common goals, cooperation, and institutional support are established contact conditions, while repetition and conflict mediation are proposed here as design supports [5].

Friction is not the goal. Rosa describes resonance as a relation in which a person is affected by something beyond the self, responds in their own voice, and is transformed through an encounter whose outcome cannot be fully controlled [6]. A child shapes a wet-sand channel; it collapses; a peer suggests another route; the child objects, revises, and retries. The episode may resonate if the world feels responsive. Friction never guarantees resonance: overwhelming risk, coerced contact, or scripted programmes can produce withdrawal. Cities cannot engineer resonance, but they can distribute its enabling conditions more equitably.

Childhood play should not be presented as a deterministic route to socially competent adulthood. Social-brain networks and sensitivity to peers continue to develop through childhood and adolescence [7, 8]. Repeated embodied play offers practice in joint attention, turn-taking, interpreting ambiguous cues, tolerating disagreement, and repairing relationships. If personalised digital or AI-mediated environments chronically displace such encounters, children may have fewer opportunities to rehearse these capacities. This is a plausible risk, not evidence that limited resonance causes neural deficits or that play guarantees adult relational competence. Resonance is used here as a normative description of responsive world-relations, not as a neuroscientific mechanism.

3. The 15-Minute Play City

The conventional 15-minute city places daily functions within a short walk or cycle [9]. For children, fifteen minutes means repeated accessibility shaped by age, ability, sensory needs, traffic, and assistance. A facility is not genuinely near if a child cannot understand the route, cross safely, stay without compulsory consumption, and return home. Proximity matters when children can repeatedly encounter, respond to, and help shape a place.

Evidence from Guilin shows why proximity alone is insufficient. Using service-facility weights derived from 188 valid child questionnaires, the study assessed 621 neighbourhoods; 85.19% received poor or very poor walkability ratings for children's health-related services [10]. Although not generalisable, the finding highlights routes, connectivity, and children's perspectives. The UN Committee's General Comment on Article 31 frames the provision of play conditions as a public obligation [11].

The model has two infrastructural layers and a cross-cutting digital principle (Fig. 1). Its spatial network links the Pocket Playground, Play Corridors, and Hub Playground. Operations add time-sharing, play-sufficiency assessment, and co-production through mapping, making, repair, and rule-setting. Both layers are necessary: a Pocket Playground without a safe Play Corridor is inaccessible; a locked schoolyard is not neighbourhood infrastructure; and co-presence without a shared task may not become relationship.

Experience before assistance is not a rule for sequencing AI within play. It is an infrastructural priority: children need sustained bodily exploration, reciprocal encounter, conflict, repair, and opportunities to form responses of their own before AI becomes a default companion, interpreter, or experiential proxy. The present concern is displacement from public life. Ambient wearable and augmented-reality systems may eventually mediate presence itself; this concern remains prospective. Communication aids, sensory previews, and accessible wayfinding belong on the enabling side because they make primary experience possible for disabled and neurodivergent children.

Child-centred but not child-exclusive, recurring use can turn strangers into familiar faces, give adolescents legitimate non-commercial public presence, and invite adults and older residents into care and stewardship. These outcomes are not guaranteed, but repetition can support agency, reciprocity, trust, and place attachment across the life course.

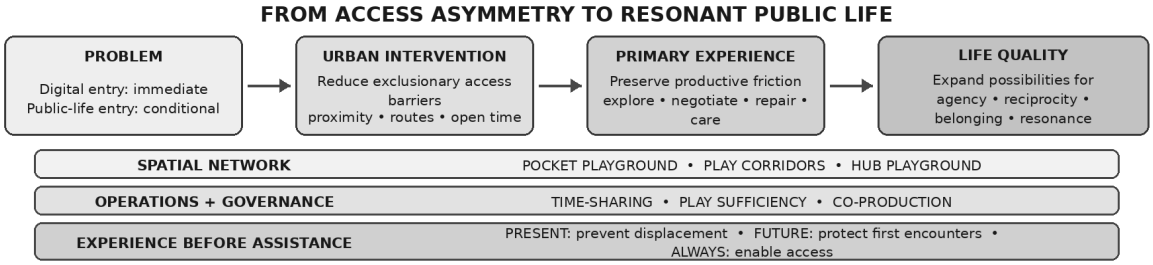


Fig. 1. The 15-Minute Play City lowers exclusionary access barriers while preserving productive friction that can enable resonance.

4. From Access Asymmetry to Everyday Public Life

Digital environments are not the urban problem in themselves. The asymmetry is that digital environments are immediate, personalised, and continuously available, while access to embodied public life often depends on parental transport, scheduled programmes, payment, safe routes, and open facilities. The goal is not to make public life frictionless: manageable physical and social resistance has developmental and civic value. The city should instead remove exclusionary friction at the threshold while preserving productive friction within experience. Primary encounter can then become an ordinary, reliable option rather than a parent-organised exception.

Implementation begins with a child-inclusive audit of distance, traffic, opening times, cost, sensory barriers, and exclusion. Germany can build on Spielleitplanung [12] and extend it through time-sharing of school grounds. The 2025 position statement calls for equitable access to active outdoor play across the life course [13]; in car-dependent North American contexts, this model translates that goal into connected Pocket Playground and Hub Playground sites, safe Play Corridors, shared-use agreements, shade, and inclusive access. Graded challenge, manipulable materials, retreat, and shared work surfaces support participation without scripting outcomes. Public funding, maintenance, and decision power are essential: co-production must extend public responsibility, not replace it with unpaid labour.

5. Limits and Conclusion

This proposition does not establish that limited resonance causes neural deficits or predicts adult relationships, or that civic friction reduces loneliness. Autonomy and risk vary by age, disability, culture, traffic, and resources; forced contact may reproduce exclusion. Pilots should evaluate independent mobility, self-initiated play, repeated contact, equity, and participant voice. Ambient AR and AI require separate study.

Digital restrictions alone cannot restore displaced experience. The 15-Minute Play City is not a city without AI. It seeks to secure public conditions for bodily encounter before digital proxies become the default. Today, this means countering displacement from public life; in an ambient-AI future, it means protecting first encounters from continuous mediation. Cities cannot produce resonance, but they can protect its possibility.

References

1. Haidt, J.: The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness. Penguin Press, New York (2024).
2. Odgers, C.L.: The great rewiring: Is social media really behind an epidemic of teenage mental illness? Nature **628**, 29-30 (2024). doi:10.1038/d41586-024-00902-2
3. Lembke, A.: Dopamine Nation: Finding Balance in the Age of Indulgence. Dutton, New York (2021).
4. Hertz, N.: The Lonely Century: How to Restore Human Connection in a World That's Pulling Apart. Crown Currency, New York (2021).

5. Pettigrew, T.F., Tropp, L.R.: A meta-analytic test of intergroup contact theory. Journal of Personality and Social Psychology **90**(5), 751-783 (2006). doi:10.1037/0022-3514.90.5.751
6. Rosa, H.: Resonance: A Sociology of Our Relationship to the World. Polity, Cambridge (2019).
7. Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., Saxe, R.: Development of the social brain from age three to twelve years. Nature Communications **9**, 1027 (2018). doi:10.1038/s41467-018-03399-2
8. Crone, E.A., Konijn, E.A.: Media use and brain development during adolescence. Nature Communications **9**, 588 (2018). doi:10.1038/s41467-018-03126-x
9. Moreno, C., Allam, Z., Chabaud, D., Gall, C., Pratlong, F.: Introducing the 15-minute city. Smart Cities **4**(1), 93-111 (2021). doi:10.3390/smartcities4010006
10. Wang, X., Huang, J., Qin, Z., Gan, W., He, Z., Li, X.: Is the Children's 15-Minute City an Effective Framework for Enhancing Children's Health and Well-Being? Buildings **15**(2), 248 (2025). doi:10.3390/buildings15020248
11. UN Committee on the Rights of the Child: General Comment No. 17 on the right of the child to rest, leisure and play, CRC/C/GC/17 (2013).
12. Child Friendly Cities Initiative: Masterplan of Play: Spielleitplanung in Regensburg. childfriendlycities.org/masterplan-play, accessed 3 August 2026.
13. Lee, E.-Y., de Lannoy, L., Kim, Y.-B., et al.: 2025 position statement on active outdoor play. Int. J. Behav. Nutr. Phys. Act. **22**, 117 (2025). doi:10.1186/s12966-025-01813-9

Hazardous Material Management in German Building Renovation and Demolition Projects

Jin Hyun Park[1]

[1]VeKNI e.V., Kneippstraße 5, 30853 Langenhagen, Germany

Abstract. Existing buildings in Germany may contain various hazardous materials, including asbestos, polychlorinated biphenyls (PCB), polycyclic aromatic hydrocarbons (PAK), and artificial mineral fibres (KMF), depending on their construction period, materials, and renovation history. During renovation and demolition, these materials present not only occupational health and environmental concerns but also significant risks to project cost, schedule, and coordination. This paper examines hazardous material management in German building renovation and demolition projects from the perspective of *Projektsteuerung*, with particular attention to the relationship between preliminary investigation, remediation planning, and project risk management. Two practical cases are presented. The first case demonstrates how the unexpected identification of asbestos during the planning stage expanded the required scope of work and led to a reassessment of project cost, feasibility, and continuation. The second case illustrates how early hazardous material investigation and the integration of remediation works into cost estimates, tender documents, and the construction schedule enabled more proactive management of project uncertainties. The cases indicate that hazardous material investigation should be regarded not merely as a technical or occupational-safety procedure, but as an early-stage project management and decision-support tool. Based on these findings, the paper discusses implications for integrating hazardous material risks into cost and schedule management in renovation and demolition projects involving ageing buildings in Korea

Keywords: Hazardous Material Management; Building Renovation; Demolition; Projektsteuerung; Project Risk Management; Asbestos

1. Introduction

In Germany, the importance of utilizing and renovating existing buildings has increased in parallel with the transition toward climate-neutral and circular construction. The continued use and renovation of existing buildings can reduce construction waste and embodied carbon associated with demolition and new construction, while also contributing to the preservation of urban landscapes and historically significant built environments.

At the same time, existing buildings in Germany may contain a range of hazardous materials depending on their construction period and previous renovation history. These include asbestos, polychlorinated biphenyls (PCB), polycyclic aromatic hydrocarbons (PAK), artificial mineral fibres (KMF), and various heavy metals. Although the risks associated with these materials may remain limited during normal building use, cutting, dismantling, or removing contaminated building components during renovation or demolition can create significant health risks for workers and potential impacts on the surrounding environment. In 2024, approximately 23.1 million tonnes of hazardous waste were generated in Germany, of which construction and demolition waste accounted for approximately 8.9 million tonnes, or 38.8%, representing the largest single share.[1] Hazardous material management in renovation and demolition projects is therefore relevant not only to occupational safety, but also to waste and resource management.

In response, hazardous material management (Schadstoffmanagement) in German renovation and demolition projects typically involves the investigation of potentially contaminated building materials prior to construction and their removal, transport, and disposal in accordance with applicable legislation and technical regulations. From a project-management perspective, the discovery of hazardous materials can affect not only the technical complexity of remediation work, but also additional construction costs, project duration, and subsequent construction activities. It is therefore important to identify and manage such risks at an early stage of the project.

This study examines the main procedures of hazardous material management through practical examples from German renovation and demolition projects and analyses the effects of hazardous material investigation and remediation on project cost and schedule from the perspective of Projektsteuerung, the German approach to client-side project management. The study aims to derive implications for the early management of cost

and schedule risks associated with hazardous materials in renovation and demolition projects involving ageing buildings in Korea.

2. Hazardous Material Management in German Renovation and Demolition Projects

Existing buildings in Germany may contain various hazardous substances depending on their construction period, the building materials used, and subsequent renovation activities. Asbestos was widely used in German construction products, particularly during the 1960s to 1980s, including fibre-cement products, insulation, plaster, adhesives, fillers, and coatings. The general ban on the manufacture and use of asbestos and asbestos-containing materials entered into force in Germany on 31 October 1993.[2] However, because buildings are often altered or renovated several times over their lifetime, the original construction date alone is not sufficient to determine whether hazardous materials are present. [2,3]

Prior to renovation or demolition, available information such as the construction date, existing drawings, and previous renovation records is reviewed. Where necessary, suspected building materials are sampled and analysed. [3,4] This preliminary investigation (*Erkundung*) provides the basis for identifying hazardous materials, selecting appropriate work procedures and protective measures, separating contaminated waste streams, estimating costs, and reducing the risk of delays during construction [3].

Table 1. Common hazardous materials in existing buildings and their risks during renovation.

Hazardous material	Typical locations in buildings	Main risks during renovation
Asbestos	Fibre-cement products, flooring and adhesives, plaster, fillers, insulation	Release of asbestos fibres during cutting, drilling, dismantling, or removal
PCB	Joint sealants, paints, coatings	Contamination of indoor air and secondary contamination of adjacent building materials
PAK	Tar-based adhesives, waterproofing materials, roofing materials	Exposure during dismantling and handling of contaminated materials and waste
KMF	Thermal and acoustic insulation	Fibre release and worker exposure during dismantling

Hazardous material management can generally be understood as a sequence of four stages: **Investigation (Erkundung) → Remediation planning (Planung) → Hazardous material remediation (Sanierung) → Clearance and disposal (Freigabe / Entsorgung).** Asbestos-related demolition, remediation, and maintenance works are subject to the German *Gefahrstoffverordnung* and TRGS 519 [5], while activities involving old mineral wool are regulated by TRGS 521 [6].

3. Case Analysis from a Projektsteuerung Perspective

In existing buildings, some areas such as wall cavities, spaces beneath floor finishes, or building-services zones cannot always be fully investigated before construction. Consequently, it is often difficult to determine the exact location and quantity of all hazardous materials during the planning stage. Project managers therefore need to consider not only the remediation costs identified through preliminary investigations, but also the possibility of additional discoveries as a cost and schedule risk. Coordination is also required among the client, designers, hazardous material consultants (*Schadstoffgutachter*), remediation contractors, and subsequent trades.

Case A. Project Reassessment Following the Discovery of Hazardous Materials

A kitchen renovation was planned in a listed historic building in Germany to improve hygienic conditions. Although the building had originally been constructed before asbestos became widely used as a building material, a hazardous material investigation was carried out during the planning stage.

The investigation identified asbestos in the existing wall plaster. The original scope of work had been limited primarily to the replacement of wall tiles. However, the required asbestos remediation expanded the scope beyond the wall finishes to include adjacent building-services components, including existing ventilation installations. As a result, the estimated remediation cost and project duration increased, requiring a reassessment of the overall project budget and economic feasibility. The project was subsequently put on hold during the planning stage.

This case demonstrates that the original construction date alone cannot reliably exclude the presence of hazardous materials. It also shows that the findings of a preliminary hazardous material investigation can influence not only the technical method of construction, but also early-stage decisions regarding whether a project should proceed.

Case B. Proactive Hazardous Material Risk Management through Preliminary Investigation

In another demolition project involving an existing building in Germany, a preliminary hazardous material investigation was conducted together with a specialist consultant before the start of the main works. The investigation identified the location and anticipated remediation scope of asbestos, PAK, PCB, and KMF. Based on these findings, the expected quantities of contaminated materials and waste were estimated and incorporated into cost calculations and tender documentation (*Leistungsverzeichnis*, LV).

Hazardous material remediation (*Schadstoffsanierung*) was also scheduled as a separate activity preceding the general demolition works. The required remediation period and the planned start of subsequent demolition activities were incorporated into the overall construction schedule. Areas that could not be fully assessed during the investigation stage were treated as residual project risks.

This approach did not eliminate the presence of hazardous materials or all uncertainty associated with them. However, it made it possible to anticipate and manage a significant proportion of the associated cost and schedule risks from an early stage of the project.

Table 2. Comparison of Project Management Responses to Hazardous Material Risks in Cases A and B

Category	Case A: Project reassessment / suspension	Case B: Proactive management
Initial investigat ion	Hazardous material investigation during planning	Detailed investigation at an early stage
Main finding	Unexpected asbestos contamination	Identification of major hazardous materials and remediation scope
Cost impact	Increased remediation cost and reassessment of overall project budget	Anticipated remediation cost incorporated into project budget
Schedule impact	Reassessment of project and programme	Remediation incorporated as a preceding activity
Projektste uerung response	Reassessment of economic feasibility and project continuation	Proactive management of cost, schedule, and interfaces
Outcome	Project put on hold	Subsequent demolition works proceeded

Although the two cases resulted in different project outcomes, both demonstrate the importance of investigating hazardous materials at an early stage. In particular, the suspension of Case A should not necessarily be interpreted as a failure of project management. Rather, the identification of previously unknown remediation costs before construction allowed the project team to reassess the economic feasibility of the project before committing further resources. From this perspective, the investigation functioned as a project risk-management and decision-support tool.

4. Implications for Korea and Conclusion

The German cases examined in this study demonstrate that hazardous materials in existing buildings should not be regarded solely as a technical issue related to demolition and remediation. They can constitute significant project risks affecting cost, schedule, and, in some cases, the decision of whether a project should proceed. Since a preliminary investigation cannot always identify all hazardous materials within an existing building, project management should address not only the materials already identified but also the possibility of additional discoveries during construction.

Several implications can be derived for renovation and demolition projects involving ageing buildings in Korea. First, hazardous material investigations should be considered not only as occupational-safety procedures immediately preceding demolition, but also as early-stage project risk assessments. Second, anticipated remediation costs and durations should be integrated into the overall project budget and construction schedule. Third, the possibility of additional hazardous material discoveries should be recognised as a source of cost and schedule uncertainty and incorporated into project risk-management strategies.

Hazardous material management should therefore be understood not only as a matter of occupational safety and environmental protection, but also as a project-management tool supporting early decision-making, cost planning, and schedule management in existing-building projects. The German cases suggest that renovation and demolition projects involving ageing buildings in Korea may benefit from integrating hazardous material management into the project-management framework from an earlier stage.

References

1. Statistisches Bundesamt (Destatis): *23,1 Millionen Tonnen gefährliche Abfälle in Deutschland im Jahr 2024*. Pressemitteilung Nr. 260, 22 July 2026.
2. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA): Informationsplattform Asbest – In welchen Fällen muss mit asbesthaltigen Bauprodukten gerechnet werden? BAuA.
3. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA): Informationsplattform Asbest – Planung und Vorbereitung: Erkundung. BAuA.
4. Bundesministerium der Justiz: Verordnung zum Schutz vor Gefahrstoffen (Gefahrstoffverordnung – GefStoffV), § 5a: Besondere Mitwirkungs- und Informationspflichten für Veranlasser von Tätigkeiten an baulichen oder technischen Anlagen.
5. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA): TRGS 519 – Asbest: Abbruch-, Sanierungs- oder Instandhaltungsarbeiten. Ausgabe Januar 2014; zuletzt geändert und ergänzt: GMBl 2025, S. 144–145, 28 February 2025.
6. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA): *TRGS 521 – Tätigkeiten mit alter Mineralwolle*. Ausgabe Mai 2026, GMBl 2026, S. 300–308.

Das Badhaus in Aachen

Minkyu Kim

Chair of Building Typologies and Design Basics,
Faculty of Architecture, RWTH Aachen University, Aachen, Germany

Abstract. This master's thesis in architecture, "Das Badhaus in Aachen," was submitted at the Chair of Gebäudelehre und Grundlagen des Entwerfens and proposes a new bathing architecture for the historic Kaiserbad site in Aachen. The work unfolds in three parts. Badekultur surveys the history of bathing culture from Greco-Roman antiquity through the medieval period to modern hygiene reforms, extending comparatively to bathing traditions in Japan and the Islamic world. Fragment draws on Aldo Rossi's notion of the city as a collective memory of accumulated fragments, cataloguing thirteen architectural and atmospheric elements—from spatial typologies like the apodyterium and natatio to sensory elements such as steam and water spouts—drawn from Roman bath typology and the site's own history. Entwurf translates these fragments into a design proposal, "Kaiserbad 6.0," reinterpreting the site's five historical building phases through plans, elevations, and sections, offering a method for reconstructing urban memory through layered reinterpretation of historical fragments.

Keywords: Bathing Culture, Architectural Fragments, Urban Memory

1. Introduction

Aachen has been a place of bathing for nearly two thousand years. Since antiquity, the site has hosted a continuous succession of bathing establishments—five distinct buildings, each replacing the last as needs and attitudes evolved. The most recent, the Kaiserbad (1995, Kasper & Klever), introduced an interior courtyard that remains largely unused and disconnected from the historic public squares framing the building.

This thesis takes that underused courtyard as its point of departure, asking how a new architectural intervention might engage the accumulated memory of bathing itself.

The thesis unfolds in three parts: a historical account of bathing culture (*Badekultur*), from Greek and Roman typologies through the medieval bathhouse to modern privatization, extended to Japan and the Islamic world; a fragment-based method (*Fragment*), adapting Aldo Rossi's notion of the city as a memory of accumulated fragments drawn from Roman typology and the Kaiserbad's own history; and the resulting design proposal (*Entwurf*), Kaiserbad 6.0, synthesized through urban analysis into plans, elevations, and sections.

2. Historical Background: The Culture of Bathing (*Badekultur*)

2.1 Antiquity to the Middle Ages

In Greco-Roman antiquity, bathing was far more than an act of physical hygiene—it bound together healthcare, leisure, cultural activity, and social exchange into daily life. In Greece, bathing facilities were tied to the gymnasion and palaestra, serving as sites for exercise, discussion, and communal encounter. In Rome, this evolved into the thermae: vast complexes combining baths with steam rooms, gardens, libraries, and dining halls, functioning as central institutions of urban life. This social character persisted, in modified form, into the medieval period; despite declining cultural standing under Christian moral attitudes, bathhouses continued to serve as sites of commercial exchange and conviviality, often combining hygienic functions with catering, music, and entertainment.

2.2 Transformation and Privatization

This social character began to erode with the spread of syphilis in the late sixteenth century, prompting medical advice against communal bathing. Flowing water was increasingly favored over communal pools, and by the nineteenth century, growing emphasis on personal hygiene drove the gradual privatization of bathing—advanced notably by Oskar Lassar's 1874 Association for People's Baths in Berlin. By the twentieth century, private bathrooms had become standard, and communal bathing culture lost its former urban significance.

2.3 Comparative Perspectives: Japan and the Islamic World

Two further traditions extend this history beyond Europe. In Japan, cleanliness is a pervasive cultural value; bathing water is customarily far hotter, and washing before entering the tub ensures the bath itself serves social and restorative purposes—linked historically to Buddhist notions of purity and expressed today in the onsen and sentō. In the Islamic world, bathing is intertwined with religious practice, notably the ritual ablution (*wudūʾ*) preceding prayer. The public ḥammām, echoing the Roman thermae's tripartite spatial sequence at reduced scale, historically served important social functions, particularly for women.

2.4 Significance for the Thesis

Across these contexts, bathing consistently exceeds its hygienic function to become a vehicle for social life, ritual, and collective memory—a trajectory that provides the historical and typological grounding for the fragment-based method developed in the following chapter.

3. Method: Fragment-Based Analysis (*Fragment*)

3.1 Theoretical Framework

"The city is a collective memory, a continuous accumulation of fragments from the past," wrote Aldo Rossi. Following this understanding, the thesis approaches the Kaiserbad not as a single, self-contained building to be designed anew, but as a site layered with accumulated fragments of memory—architectural, historical, and cultural—to be identified, collected, and reassembled into new spatial expression. Rather than proceeding through quantitative analysis, the method relies on interpretive selection, collecting traces of bathing culture that remain legible in typology, atmosphere, and urban memory.

3.2 Four Categories of Fragments

The collected fragments are organized into four categories: elements of the Roman bathhouse—its spatial typology, standardized bathing rituals, and social activities; fragments drawn from the broader history of bathing culture connected to the city; fragments found directly within the Kaiserbad itself, which the thesis identifies as having undergone four historical transformations, with the present design constituting a "fifth transformation"; and atmosphere—the specific mood and sensory quality of bath architecture, which the design seeks to reflect rather than merely document.

3.3 Cataloguing the Fragments

Thirteen individual fragments are identified and catalogued, each paired with an axonometric drawing and a corresponding historical image. These range from architectural elements—the old Kaiserbad's façade, historical bathtubs, the *Fürstenbad* (princely bath), the natatio, the nymphaeum, the xystus, and the apodyterium—to sensory and material fragments such as thermal water, steam, water spouts, and bath furniture, as well as social practices like massage therapy and the ephebeum.

3.4 Significance for the Design

Together, these fragments constitute a working typological vocabulary rather than a fixed template. The space is treated as a fabric of overlapping fragments from different historical layers, allowing the site's urban memory and identity to be reinterpreted rather than replicated—forming the conceptual basis for the design proposal in the following chapter.

Fig. 1. Façade of the Kaiserbad, destroyed after World War II, and its reinterpretation in the new facade design.

4. Design Proposal (*Entwurf*): Kaiserbad 6.0

4.1 Site and Urban Analysis

The design begins with the site's layered history: the plans of all five historical building phases—Kaiserbad 1.0 through 5.0—are overlaid onto the current urban fabric to reveal how each successive structure positioned itself within the city. Further diagrams examine the site's position between open and enclosed urban space, its pedestrian accessibility, and its role within the city's broader centrality structure. These analyses reveal the Kaiserbad as situated at a transitional threshold—between an accessible rear and a vehicle-dominated front—within a zone of high urban transformative potential.

4.2 Design Response: Kaiserbad 6.0

Informed by these analyses and the fragment inventory developed previously, the design proposal—Kaiserbad 6.0—reactivates the underused interior courtyard of the 1995 building, reconnecting it to the historic public squares at the building's front and rear. Architectural fragments such as the apodyterium and natatio, the atmospheric qualities of steam and water, and elements from the site's own history are integrated into a contemporary spatial and material language, expressed in the building's angled roofscape, arched openings, and layered interior circulation.

4.3 Architectural Documentation

The proposal is documented through floor plans (basement to upper floors), two elevations, and four sections (a-a′ through d-d′), concluding with a collage that superimposes the design onto historical bathing imagery, asserting continuity between the new architecture and the memory it draws upon.

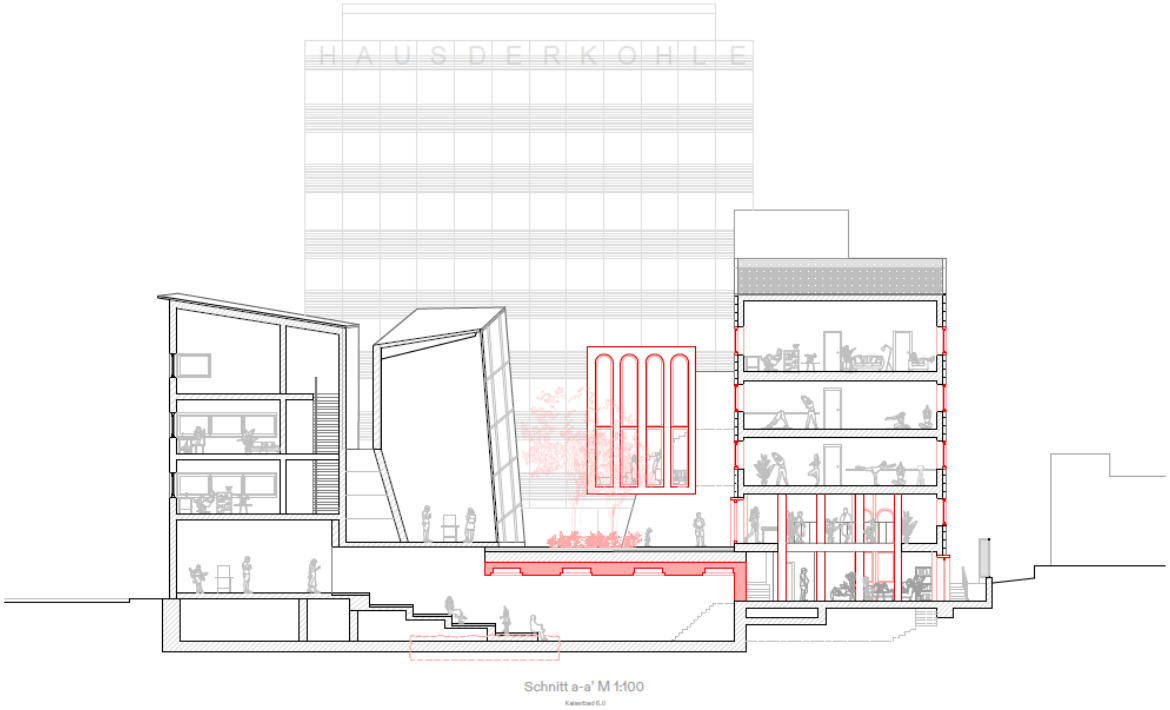


Fig. 2. Section through the interior courtyard of Kaiserbad 6.0 and the surrounding buildings.

5. Conclusions

This thesis reconsiders the underused courtyard of the 1995 Kaiserbad by developing a fragment-based design method, drawing on Aldo Rossi's notion of the city as accumulated memory. Through historical research into bathing culture and the cataloguing of thirteen architectural fragments, it produces a design proposal, Kaiserbad 6.0, that reactivates the site's connection to its historic public squares. The thesis ultimately positions architecture as an act of urban remembrance—reinterpreting a site's layered history rather than replicating or ignoring it.

References

1. Lamberto Amistadti, Valter Balduci, Tomasz Bradecki, Enrico Prandi, Uwe Schröder.: ArcheA Mapping the City
2.Fikret Yegül.: Bath and the Bathing in Classcal Antiquitty:
3. Rolf Bothe (HRSG.).: Kurstädte in Deutschland zur geschichte einer Baugattung:
4. Francoise De Bonneville.: Das Buch vom BAD:
5. Ernst Kasper, Klaus Klever.: Das Kaiserbad Der Neubau und sein historischer Hintergrund:
6. Heinz Cüppers, Walter Sage, Gustl Strunk-Lichtenberg, Erich Meuthen, Leo Hugot, Joachim Kramer, Matthias Untermann, Walter Sölter und Dorothea Haupt.: AqvaeGranni
7. Stadt Aachen.: Bad Aachen: Seine Quellen und das Kaiserbad:
8. Heinrich Klotz.: Moderne und Postmoderne:

Constructing a European Smart City: Digital Infrastructure, Urban Services, and Local Implementation in Berlin

Eunyoung Jang

Department of Social Sciences, University of Basel, Basel, Switzerland

e.jang@unibas.ch

Abstract. This paper examines how Berlin constructs a European smart-city model through the interaction of digital infrastructure, public services, and local governance. Rather than treating the smart city as a purely technological project, it conceptualizes digital systems as elements of the built environment that reshape service delivery, mobility, energy systems, and everyday spatial practices. The study combines qualitative policy-document analysis with Semantic Network Analysis of Berlin-related digital-policy and urban-policy discourse from 2023 to early2026. The findings identify three recurring policy modules: digital infrastructure, digitally integrated urban services, and data governance and regulation. Although these modules converge with broader European smart-city agendas, their implementation remains locally specific. Berlin operates as a critical laboratory in which global technological ambition encounters strong data-protection norms, decentralized administration, public-sector responsibility, and contested urban-development priorities. The cases of EUREF-Campus, Berlin TXL, and Gewobag smart housing illustrate distinct implementation pathways across energy, redevelopment, and housing. This study contributes to smart-city and urban-governance research by proposing a framework of modular convergence and contextual implementation that explains how globally circulating digital policy models are institutionally and spatially reconfigured in a post-industrial European capital.

Keywords: Smart City, Berlin, European Model, Digital Infrastructure, Urban Governance, Built Environment, Data Governance, Local Implementation

1. Introduction

Smart-city strategies have become a widely circulated policy response to urban challenges such as traffic congestion, housing pressure, energy transition, and climate adaptation. Across cities, similar technological components recur: broadband networks, urban data platforms, digitally integrated public services, smart grids, and mobility applications. Yet their circulation does not produce a uniform urban model. Technologies are introduced through existing administrative systems, planning cultures, procurement rules, and public expectations; consequently, the same digital module can generate different institutional and spatial outcomes.

Berlin offers a particularly revealing case because global ambitions for data-driven urban development encounter an institutional environment characterized by strong data-protection norms, decentralized administrative structures, public-sector accountability, and a diverse field of municipal, private, and research actors. Berlin therefore functions as a critical laboratory in which technological ambition is negotiated through European regulation and local governance. The central question is: how does Berlin translate globally shared smart-city paradigms into locally specific infrastructures, services, and governance arrangements? The paper argues that Berlin's model combines convergence in policy design with contextual divergence inimplementation.

2. Theoretical Framework and Method

The analysis draws on the concepts of institutional convergence and local translation. World-society and institutional-isomorphism approaches explain why governments increasingly adopt comparable policy categories and organizational models [1, 2]. In smart-city policy, convergence can be observed in the repetition of common modules, including digital connectivity, platform-based services, data governance, and technology-enabled sustainability. Yet these modules are not transferred as complete and rigid systems. They are reconstructed through local institutions, legal frameworks, and urban-development priorities.

This paper also treats digital technology as urban infrastructure rather than as a separate virtual layer. Networked systems structure access to energy, mobility, housing, and public services and therefore participate in the production of urban space [3]. From this perspective, smart-city development is a process of infrastructuration: digital systems become embedded in the built environment and influence how services are organized, where resources are allocated, and how citizens interact with public institutions.

Empirically, the study combines qualitative policy-document analysis and Semantic Network Analysis (SNA). The source material includes Berlin Senate policy documents, institutional project materials, and media discourse on digital and urban policy from 2023 to early 2026. Policy categories and recurring relationships were coded and interpreted with MAXQDA. The source selection balances top-down policy documentation with publicly circulated media discourse, capturing both official strategic priorities and their broader interpretation in the urban public sphere. This triangulation provides a more comprehensive view of how smart-city concepts move from policy formulation to local implementation. Because this short paper summarizes a broader research project, the SNA is used primarily to identify dominant thematic clusters and temporal tendencies rather than to present a complete set of network statistics.

3. Findings and Discussion

The analysis identifies three central policy modules in Berlin's digital urban agenda. The first is digital infrastructure, including broadband expansion, gigabit connectivity, smart grids, sensor systems, and mobility-data ecosystems. This module establishes the technical foundation on which other smart-city applications depend. The second is urban services, especially digital public administration, multimodal mobility, energy management, and digitally supported housing services. Here, technology is presented as a means of increasing accessibility, coordination, and operational efficiency. The third module is governance and regulation, covering data sovereignty, platform coordination, privacy, interoperability, public accountability, and compliance with German and European regulatory standards.

These modules demonstrate convergence with wider European and international smart-city agendas. Berlin does not invent a unique technological vocabulary; it adopts a shared toolkit also visible in other metropolitan strategies. The local specificity of the Berlin model lies instead in the way these components are assembled. Implementation is negotiated across Senate departments, district administrations, municipal companies, research institutions, and private partners. Legal requirements, public procurement, data protection, and civic expectations can slow implementation, but they also make institutional responsibility and public legitimacy central elements of digital transformation.

A temporal reorientation is visible in the policy discourse. During 2023-2024, Berlin's agenda concentrated primarily on connectivity, pilot projects, and the expansion of technical infrastructure, often at the district, campus, or building scale. From 2025 onward, however, the emphasis increasingly moved toward regulatory alignment, institutional coordination, and integrated data governance. Consequently, this development should not be read as a replacement of infrastructure-oriented policy, but as its gradual incorporation into a broader governance framework.

Three local cases illustrate how abstract digital modules become embedded in Berlin's built environment. EUREF-Campus in Schöneberg integrates smart grids, electric-vehicle charging, sustainable energy systems, and mobility experiments within a cooperative ecosystem of firms and research organizations. Its significance lies not only in the technologies deployed, but also in the governance model that coordinates energy transition, innovation, and urban development in a defined site.

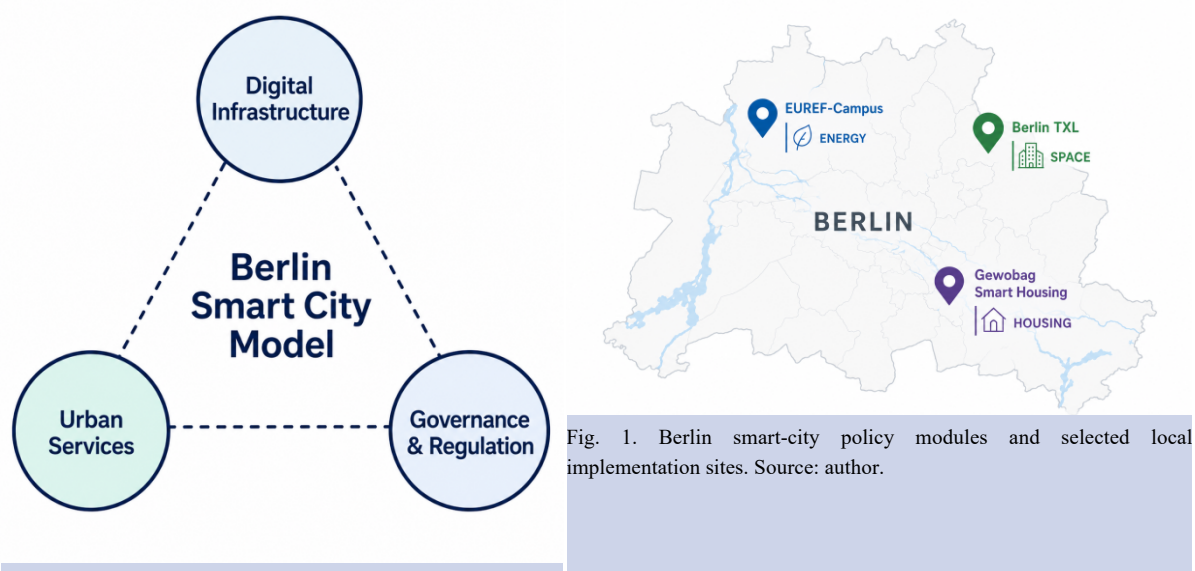


Fig. 1. Berlin smart-city policy modules and selected local implementation sites. Source: author.

Berlin TXL translates smart-city policy into the regeneration of the former Tegel Airport. The Urban Tech Republic combines sustainable buildings, energy systems, mobility technologies, research facilities, and long-term spatial planning. This case demonstrates how digital infrastructure becomes part of large-scale urban redevelopment rather than an add-on to an already completed urban form. The project connects technological experimentation with land-use transformation and economic-development objectives.

Gewobag's digital and smart-housing initiatives provide a third implementation pathway. Smart-home systems, heating management, charging infrastructure, and mobility hubs are linked to the responsibilities of a municipal housing company and to the provision of affordable housing. The case highlights the social dimension of smart urbanism: digitalization becomes relevant when it improves everyday service delivery and remains accessible to residents rather than functioning only as a premium technological product.

Taken together, the three cases demonstrate that Berlin's smart-city development does not follow a single implementation pathway. EUREF-Campus represents an innovation-oriented energy and mobility ecosystem driven by collaboration among business, research, and infrastructure providers. Berlin TXL incorporates digital technologies into the large-scale redevelopment of a former airport, linking experimentation with long-term spatial planning. Gewobag's initiatives instead embed digital systems within publicly oriented housing provision and everyday residential services. Despite their sectoral differences, all three cases face a common governance challenge: digital technologies must be coordinated across multiple actors while remaining compatible with regulatory standards, spatial-planning objectives, and public-value considerations. Their comparison therefore shows how shared modules are translated differently according to site, ownership structure, policy domain, and institutional responsibility.

4. Conclusions

This study examined how Berlin constructs a smart-city model through globally circulating policy modules and locally embedded implementation. The analysis identified convergence around digital infrastructure, integrated urban services, and data governance, while also revealing a gradual movement from infrastructure-led experimentation toward regulatory and institutional integration. Berlin's strong regulatory environment and fragmented but collaborative governance structure do not simply obstruct technological innovation; they actively reshape its form and public legitimacy.

The cases of EUREF-Campus, Berlin TXL, and Gewobag demonstrate that digital transformation becomes urban transformation only when technologies are integrated into energy systems, mobility, redevelopment, housing, and everyday service provision. A framework of modular convergence and contextual implementation thus offers a critical lens for understanding why cities may adopt similar smart-city vocabularies while producing different institutional and spatial outcomes. For planning and policy, the main implication is that successful smart urbanism depends not only on technological capacity, but also on spatial integration, governance capacity, regulatory accountability, and public accessibility.

References

1. DiMaggio, P.J., Powell, W.W.: The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review 48(2), 147-160 (1983).
2. Meyer, J.W., Boli, J., Thomas, G.M., Ramirez, F.O.: World society and the nation-state. American Journal of Sociology 103(1), 144-181 (1997).
3. Graham, S., Marvin, S.: Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition. Routledge, London (2001).
4. BBSR: Smart Cities und digitale Stadtentwicklung: Modellprojekte, Governance und kommunale Umsetzung. Bundesinstitutfuer Bau-, Stadt- und Raumforschung, Bonn (2024).
5. Land Berlin: Senate policy documents and press releases on digitalization, mobility, urban development, and smart-city governance, 2023-2026. Berlin Senate Administration (2023-2026).
6. Jang, E.: The Alignment and Modeling of Smart City Models between Berlin and Seoul. Unpublished research report, Humboldt-Universitaetzu Berlin (n.d.). Accessed 23 Aug 2026.
7. EUREF AG: EUREF-Campus Berlin: Energy transition and sustainable urban innovation. EUREF-Campus Berlin (n.d.). Accessed 23 Aug 2026.
8. Tegel Projekt GmbH: Berlin TXL - The Urban Tech Republic. Tegel Projekt GmbH, Berlin (n.d.). Accessed 23 Aug 2026.
9. Gewobag: Gewobag ID: Digital innovation and smart housingprojects. Gewobag Wohnungsbau-Aktiengesellschaft Berlin (2026).

Experiences and Advice from Scientists and Engineers for the Next Generation

Aims

This session provides a platform for scientists and engineers to share their personal and professional experiences throughout their careers. Contributions may address a broad range of topics, including studying abroad, academic and research life, international collaboration, transitions into industry, research funding acquisition, and career development. By sharing practical experiences, challenges, achievements, and lessons learned, this session aims to provide valuable insights into the diverse career paths and opportunities available to scientists and engineers.

The session particularly welcomes contributions offering advice and guidance to the next generation of scientists and engineers. Authors are encouraged to discuss strategies for navigating academic and professional environments, developing international research networks, securing research funding, pursuing interdisciplinary collaborations, and making informed career decisions. Through the exchange of knowledge and experiences across generations, the session aims to foster professional growth, strengthen the scientific community, and inspire young researchers to pursue successful and meaningful careers in science and engineering.

국가전략기술과 글로벌 경력 개발 – 양자기술을 중심으로

박갑동[1]

[1] KE-QSTCC Rue de la Science 14A, 1050 Brussels, Belgium
kdpark0326@gmail.com

Abstract. 과학기술 선진국들은 산업 발전은 물론 외교/안보 주권 확보라는 이중 용도를 가진 국가전략기술 선점에 국가 역량을 집중하고 있다. 국가전략기술은 핵심기술 자체의 첨단성뿐만 아니라 소재/부품/장비 공급망, 연구 인프라, 기술 활용 및 사업화, 인력 확보 등 전주기 기술혁신의 다양한 요소들과 거대한 예산 투입을 필요로 한다. 따라서 이 기술 개발과 사업화 전 과정은 어느 한 나라가 독점하기는 불가능하기에 국제협력이 필수적이다. 개별 연구자의 관점에서는 경력 경로의 전주기 과정에서 이 사업에 참여가 가능하므로 글로벌 경력 개발에 매우 효과적이다. 본 기고에서 필자는 지금까지 산학연에서의 연구 및 사업화 경험과 최근 양자기술 국제협력 활동을 바탕으로 과학기술 후속세대를 위해 바람직한 글로벌 경력 개발이 무엇인지 대한 고민을 공유하고자 한다.

Keywords: National Strategic Technologies, Dual Uses, International Cooperation, Global Career Development

1. 서론

20세기초 물리학을 뒤흔든 30년 동안 세기의 논쟁 끝에 지금의 학문의 한 영역으로 진입한 양자역학은 핵자기공명(NMR) 현상과 이를 이용한 자기공명영상(MRI) 개발, 원자에 대한 깊은 이해와 불행한 원자탄의 등장, 레이저 개발, 반도체 및 집적회로(IC) 개발 등 1차 양자혁명의 결과물을 통해 세상에 큰 영향을 주면서 오늘에 이르고 있다. 1980년대 초부터 반도체 분야에서 무어의 법칙의 한계와 CPU 기반 디지털 계산의 한계를 자연의 기본 현상인 양자기술로 극복해보자는 일련의 아이디어가 나타났다. 결국 양자정보기술 기반의 알고리즘 개발, 초전도 양자 큐비트 기반의 양자 Annealing 개발과 IBM의 양자컴퓨터 개발 및 양자 우위(Quantum Superimacy)로 이어지면서 양자정보기술에 의한 2차 양자혁명을 초래하고 있다. 정보를 습득하고(양자 센싱), 정보를 처리하며(양자 컴퓨팅), 정보를 전달하는(양자 네트워킹) 전 과정에 양자역학의 기본 현상을 적용하면서 다시 한번 더 세상을 놀라게 한다.[1]

과학기술 선진국들은 자국의 차별화 특성과 현황 및 미래 전망을 바탕으로 양자와 같은 혁신기술을 국가전략기술로 선정하여 연구개발과 기술 사업화에 국가 역량을 집중하고 있다. 이들 국가전략기술에 대한 투자는 균형 육성을 위해 골고루 뿌려주는 방식이 아니라 선택과 집중(Selection & Focus)을 통해 육성한다. 균형 지원과 선택과 집중에 의한 육성의 비율은 나라마다 다르고, 기술 발전의 단계마다 다르다. 대부분 기초연구 이전 단계에서는 상향식(Bottom-Up) 균형 육성을, 산업적 파급효과가 입증되면 집중 육성을 하는 것이 일반적인데 최근 기술 주기가 짧아지고 기술 융합과 연구 외적인 변수로 인해 이 경계가 무너지는 사례가 늘어나고 있다. 미래 사회의 불확실성, 글로벌 경쟁 심화, 불안정한 공급망으로 인해 이들 국가전략기술들은 대부분은 산업 발전은 물론 국가 안보/외교 차원에서 글로벌 공급망에 밀접하게 의존하는 이중 용도(Dual-Use) 기술에 해당한다. 우리나라도 양자기술을 포함한 10대 기술을 국가전략기술로 선정하여 중장기 발전전략 및 정책 수립, 법/제도 제정, 사업 발굴 및 기획, 예산 확보 및 투입, 인프라 구축, 인력 확보 등 전방위로 역량과 자원을 집중하고 있다. 특히, 양자기술은 AI, 첨단 바이오와 함께 미래를 바꾸는 Game Changer 기술로 선정되어 더욱 큰 관심을 받고 있으며, 양자-AI 하이브리드 사업 추진으로 AI 기술의 한계를 극복하면서 한층 강화된 응용 능력을 보일 것으로 주목하고 있다.

이들 국가전략기술의 발전을 위해 가장 중요한 것은 인력 양성이다. 양자기술도 양자산업은 물론 공공안전에 미칠 엄청난 잠재력과 막대한 자원 투입에도 전 세계적으로 양자기술 인력은 수요에 크게 못미치고 있다.[2] 양자대학원과 같은 석·박사 과정과 박사후연구원(Post-Doc)을 거치는 전문 연구자뿐만 아니라 기존 학문과 산업 영역에서 양자 분야로 전환하는 연구자도 포함해서 다양한 전문성과 역량을 가진 인력을 필요로 한다. 이를 바탕으로 양자와 직접 관련된 연구자와 양자 전환을 시도하는 연구자들을 위해 몇가지 경험을 공유하고자 한다. 특히, 미래를 이끌어갈 과학기술 후속세대와 바람직한 연구자로서 홀로서기를 준비하는 과학기술인들에게 초점을 맞추었다.

2. 본론

과학기술이 어떻게 세상을 바꾸면서 지속 가능한 발전을 이루는지는 국가혁신시스템(National Innovation System, NIS) 개념이 잘 설명하고 있다. 생존과 발전을 위해서는 기존의 축적된 과학을 배우고 익혀서 신지식을 발견하고, 신기술을 개발하며, 이를 신사업으로 연결시켜 산업 발전과 공공 이슈를 해결함으로써 선순환을 지속해야 한다. 결국 이것은 교육, 연구, 개발, 혁신의 선순환을 말하고, 이를 누가 어떤 과정을 거쳐 실행하는지 체계적으로 정리한 것이 국가혁신시스템 이론이다. 이 이론을 양자기술에 적용하여 양자기술이 어떻게 세상을 바꾸고 미래에 활용되는지를 양자혁신시스템 개념으로 살펴보고, 바람직한 경력 개발에서 무엇이 중요한지 경험을 공유하고자 한다.[3]

<table>
<tr><th>양자과학</th><th>양자 분야</th><th>핵심 양자기술</th><th>활용(User Case)</th><th>확산</th></tr>
<tr><td rowspan="5">중첩
얽힘
측정
(붕괴)</td><td>양자 컴퓨팅
(초고속 연산)</td><td>• Qubit 개발(6+1) : 초전도, 이온트랩, 중성원자, 광자, 반도체 스핀, 고체점 결함 +마요나라
• QPU 개발 → Full-Stack System
- H/W, 제어 S/W, 활용 Algorithm
- 양자 Emulator → 양자 Annealer
- NISQ → FTQC(QEC) → 범용 양자컴퓨터</td><td>• 조합 최적화
• 미분 계산 → 예측
• 분자 시뮬레이션
• 양자기계학습(QML)</td><td rowspan="5">• 양자기술 사업화
- Spin-Off/Spin-Out
- Start-up → Scale-up

• 관련 산업 적용
- 인공지능
- 바이오/신약
- 의료/헬스케어
- 에너지/소재
- 금융
- 교통/물류
- 통신/보안
- 국방/안보</td></tr>
<tr><td>양자 통신
(초신뢰 보안)</td><td>• 양자암호(PQC, QKD, Hybrid)
• 양자 네트워크/인터넷</td><td>• 양자보안 서비스
• 양자 인터넷 서비스</td></tr>
<tr><td>양자 센싱
(초정밀 계측)</td><td>• 중력/시간/자기장/전기장/온도 센서</td><td>• 초고분해능 센싱</td></tr>
<tr><td colspan="3">소재/부품/장비 – 글로벌 공급망</td></tr>
<tr><td colspan="3">양자 인프라
• 양자 하이브리드(Q-HPC/AI), 양자 팹, 양자 파운드리, 양자 테스트베드</td></tr>
</table>

그림 1. 양자기술 사업화 전주기 프로세스

첫째, 과학기술의 사업화 전 과정을 한 눈에 볼 수 있어야 하고, 그 속에서 나의 기술은 어디에 있는지 현재 주소를 정확히 알아야 한다. 자신의 연구 분야가 좁다고 하더라도 국가전략기술은 어느 한 기술뿐만 아니라 다양한 기술을 필요로 하고, 심지어 지금까지 없는 새로운 기술을 만들어내는 거대한 프로세스다.

그림 1은 양자기술의 구조와 전주기 프로세스를 표로 정리한 것이다. 양자과학을 바탕으로 양자컴퓨팅, 양자통신, 양자센싱의 3가지 양자 분야와 해당 핵심 양자기술을 나타냈으며, 이로부터 다양한 응용 또는 User-Case를 발굴하여 양자기술 자체의 산업뿐만 아니라 더 막대한 연관 응용산업을 발전시킨다. 이를 성공적으로 수행하기 위해 기술 분류(technology Tree)에 따른 소재/부품/장비 개발과 공급, 양자 팹/파운드리/클라우드 및 테스트 등 인프라 지원도 중요하다. 이들 모든 영역에 전문 인력 양성이 함께하는 것은 당연하다. 나의 기술의 현재 주소는 물론 앞으로 어느 영역으로 이동을 할 지가 경력 경로(Career Path)로 나타나며,

어떤 경로를 갈지 선택하는 것이 경력 개발에서 가장 중요하다.

둘째, 양자기술의 전 과정을 움직이는 주체들이 누구이고, 그 속에서 나는 현재 어디에 위치하는지를 알아야 한다. 그런 다음에 나의 협력자들은 어디에 있는지를 파악하게 된다.

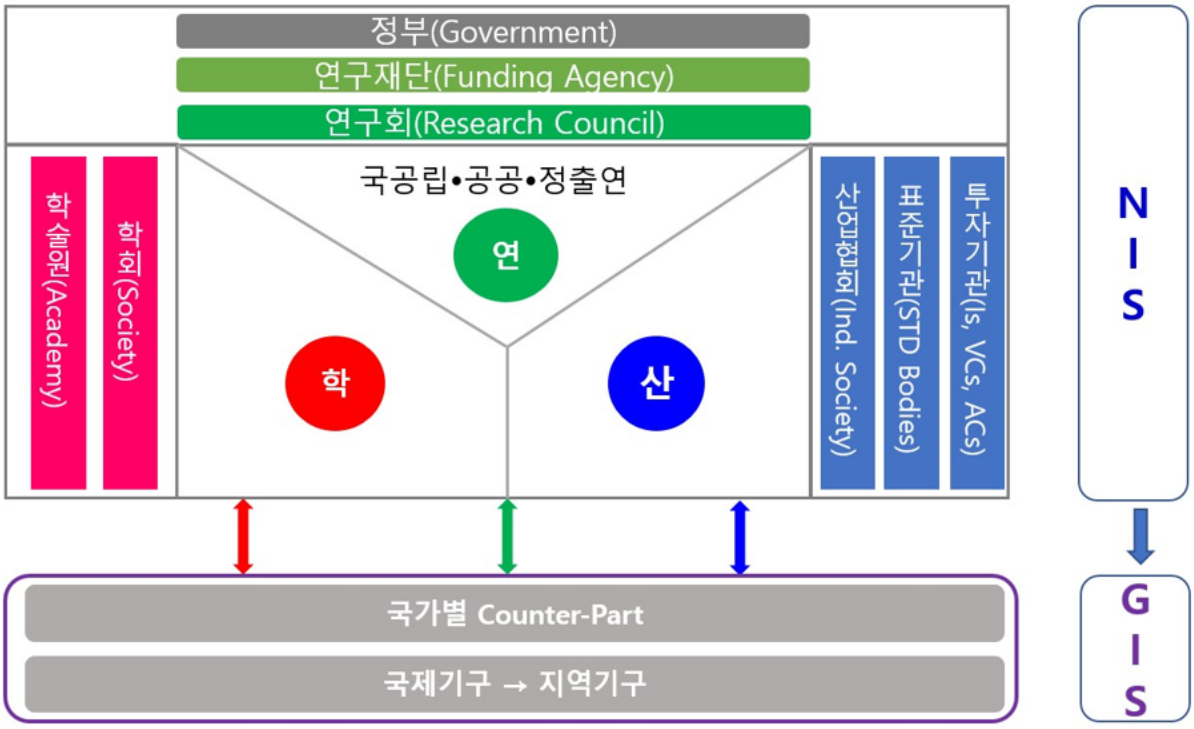


그림 2. 국가혁신시스템과 글로벌혁신시스템 – 거버넌스 중심

그림 2는 한 나라의 과학기술 혁신을 주도하는 조직/기관이 누구인지를 나타낸 것이다. 먼저 국가혁신시스템은 크게 핵심 활동(Primary Activities)과 지원 활동(Supporting Activities)으로 나눌 수 있다. 지식을 발견하고, 기술을 개발하며, 사업화로 연결하는 것이 핵심 활동이다. 이는 주로 대학연구소, 국공립연구소, 기업연구소의 연구조직과 소속 연구자들이 담당하는데, 특히 이를 산학연 삼두마차(Toika)라고 부른다. 대학은 학술원, 한림원, 전문학회와 같은 학계(Academic Societies)와 연계되는데, 주로 교육과 연구를 담당하고 최근 연구중심대학을 중심으로 기술사업화와 창업과 같은 혁신 활동도 활발히 추진한다. 국공립연구소는 과학기술부, 연구재단(Funding Agency), 연구회(Research Council)으로부터 미션과 예산을 받고, 책임과 권한을 가지고 임무를 수행한다. 물론 연구 자율성, 독립성, 전문성 보장이 중요한 것은 역사적으로나 실제 현장을 통해서도 잘 알고 있다. 기업은 영업도 중요하지만 갈수록 자체 연구도 더욱 중요하다. 최근에는 기업의 연구 역량 확대로 정부와 국공립연구소를 앞서면서 오히려 첨단 연구를 압박하고 주도하는 것을 자주 확인한다.

핵심 활동과 연계해서 전략/정책/법/제도 수립, 예산 확보 및 투자, 국책사업 발굴 및 기획, 인프라 제공 등 연구 외적인 모든 것을 제공하는 생태계 구축은 지원 활동에 해당한다. 핵심 활동도 중요하지만 지원 활동도 못지않게 중요하다는 것은 주요 선진국들의 사례를 통해 잘 알려져 있다. 연구 활동에서 연구 지원활동으로 경력 전환을 하는 것은 일상적이며, 오히려 연구 활동 경험을 한 적이 있거나 가까이서 오래 협력한 경험이 있는 연구 지원 전문가들의 성공 가능성이 높은 것은 자명하다. 그리고 연구 활동의 범위를 전 세계로 확장하여 글로벌혁신시스템(Global Innovation System, GIS)으로 연결하는 국제협력도 지원 활동의 한

부분이며, 글로벌 공급망(Global Supply Chain) 차원에서 더욱 중요성이 높아지고 있다. 한 나라의 혁신시스템과 일대일로 매칭이 되는 상대국의 파트너가 같은 구조 속에서 존재하고, 이를 국제적으로 확대하는 것이 역량과 성과를 높이는 지름길이다. 과연 나는 누구와 협력과 경쟁을 하고, 네트워킹을 통해 역량과 전문성을 높일 수 있는지 그 구조와 역학관계를 잘 알아야 한다.

셋째, 앞에서 제시한 전체 흐름과 구조를 바탕으로 어떻게 네트워킹할 것인지 연결하고, 협력하는 것이다. 어떻게 내게 필요한 정보를 얻고, 적합한 연구기관이나 파트너를 찾을 것인지가 중요하다. 먼저 한 박자 빨리 시작하는 것은 기본이다. 무조건 빨리 시작하라는 것이 아니라 전문가가 아닌 이상 적어도 정상적인 진행 단계보다 한 템포 빠르게 시작해야 겨우 흐름에 맞출 수 있다. 처음 입문하는 단계라면 3개의 플랫폼을 찾는 것이 좋다. 학계가 주도하는 학회(Academic Societies)의 학술대회(Conference), 정부와 연구재단(Funding Agency)이 주도하는 국가전략사업 기획 및 공모 과정, 기업이 주도하는 전시회/박람회(Exhibition) 또는 민간 주도 Conference가 3대 플랫폼이다. 플랫폼에는 지식/기술, 사람, 제품/서비스가 한 눈에 공개되고, 거래가 이루어진다. 즉, 양자기술 분야의 산학연관 전문가들과 해외 전문가들까지 참가한다. 그 속에서 나를 소개하고, 나의 파트너를 만나고, 새로운 아이디어를 발견하는 출발이 이루어진다. 이 플랫폼에서 나는 무엇을 팔거나 줄 수 있는지를 명확히 소개하는 것은 필수다. 그리고, 내가 원하는 것이 무엇인지와 무엇을 찾고 있는지도 함께 제시할 수 있어야 한다. 주고 받는 것을 명확히 설정할 수 있어야 다음 단계를 시작할 수 있다. 온라인으로 지식과 기술을 검색하고, 워크샵/세미나/포럼에 참석하는 것도 중요하지만 직접 얼굴을 마주하고 소통하는 것은 더욱 중요하다. 현재의 네트워킹은 수명이 있고, 늘 한계에 직면한다. 정기적인 네트워킹 채널이 없고, 활동이 미약하면 나의 기술도 죽고, 미래의 경력 개발도 약해진다.

넷째, 네트워킹을 통해 파트너십을 찾고, 경력 개발 상에서 나의 진로를 결정하는데 무엇을 참고할 것인가? 1980년대에 우수연구센터(Center of Excellence) 선정의 4가지 기준을 참고하면 도움이 된다.[4] 그 분야의 대가가 있는가? 대가는 없더라도 Dream Team을 꾸리고 있는가? 독보적인 시설/장비가 있는가? 축적된 지식/기술/데이터가 있는가? 물론 이 4가지를 다 갖춘 곳이면 더할 나위가 없겠지만 처음부터 찾기는 어렵다. 이 중에서 어느 하나라도 나의 전문 분야와 연계되어도 성장 가능성은 충분히 있고, 노력에 따라 확장할 수도 있다. 그리고, 나의 경력 개발 전 과정에서 4가지 요소는 중요한 지침과 기준점이 된다.

다섯째, 마지막이자 가장 중요한 요소인데, 네트워킹 확대는 물론 팀 내부 커뮤니케이션에서 신뢰를 형성하는 자세 또는 태도의 문제이다. 결국 오래, 깊이, 멀리 가기 위해서는 신뢰가 중요하고, 이는 나의 자세와 태도에 직결된다. 경력 개발 사다리 또는 징검다리를 건널 때마다 많은 새로운 소통과 신뢰를 필요로 한다. 쉽지 않지만 자세나 태도가 좋은 습관으로 형성되면 무엇보다 강력한 동반자가 될 수 있다. 먼저 타인에 대한 존중은 필수다. 상대방 인격과 품성에 관계없이 기본적으로 존중하는 자세가 가장 중요하다. 어렵고 불편한 상대는 전체가 아니라 일부분만 집중하면 된다. 그 다음은 항상 상대방의 의견을 경청하고, 이를 바탕으로 질문과 토론을 이어가는 것이다. 다수가 토론할 때에 경청은 쉽지 않다. 모호할 때는 다시 질문하거나 상대방의 요지를 확인하는 것이 상대방에 대한 존중과도 연결된다. 토론의 꽃은 심층 토론이다. 아무리 지식이 많고 기술이 좋아도 충분히 설명하고 또 이해되지 못하면 소용이 없다. 토론을 이끄는 사회자가 없더라도 매끄러운 진행자(Facilitator) 또는 중재자(Moderator) 역할을 할 수 있으면 좋다. 결국 상대방으로부터 충분한 이야기를 이끌어내고, 다양한 전문가들의 손길을 통해 새로운 가능성을 열어야 나에게 기회가 열리는 것이다. 그러기 위해서는 먼저 봉사정신(Sevantship)이 있어야 하고, 그래야 팀웍으로 연결된다. 직접적인 나의 성과도 중요하지만 나로 인해 동료나 파트너가 더 큰 기회를 창출하면 나의 발전에도 도움이 된다는 열린 자세를 가져야 한다. 그리고, 최종 결과를 정리해서 공유하는 것까지 완결하는 책임있는 모습을 보여줘야 한다. 눈에 보이는 단기적인 손해가 결국 장기적인 결실로 이어진다는 사실은 네트워킹 활동을 오래한 연구자들이 늘 강조하는 말이다. 누구나 알고 있는 내용이지만 실제로 실천하기는 쉽지 않다. 설문조사에서도 가장 중요하면서도 어려운 부분으로 꼽히는 것이 바로 이것이다. 결국 이 부분에는 특별한 왕도가 없다고 생각한다.

3. 결론

지금까지 국가전략기술의 하나인 양자기술 분야에서 글로벌 경력 개발을 어떻게 이어가면 좋을지 바람직한 경력 개발을 나름대로 경험을 바탕으로 제시해 보았다. 양자기술을 국가혁신시스템 기반으로 전체 혁신 프로세스와 거버넌스 구조를 분석틀로 삼아 과학기술 후속세대를 위한 바람직한 경력 개발의 핵심성공요인을 나열했다. 양자기술의 개발부터 사업화 전 과정에서 나의 기술 분야는 어디에 위치하고, 양자혁신시스템을 움직이는 주체들은 누구인지를 명확히 알아야 한다. 그리고, 이들 주체들과 연결할 수 있는 네트워킹 채널은 무엇이고, 네트워킹 채널의 파트너들과 무엇을 주고받을 것이며, 나의 다음 경력 개발을 위한 징검다리는 어떻게 선택할 것인지도 제시하였다. 마지막으로 가장 중요한 소통과 신뢰는 나의 자세와 태도 관점에서 정리하였다.

인공지능의 발전과 국제 질서의 불확실성 속에 과학기술을 이어가는 후속세대들은 어려움도 많지만 발전한 도구와 네트워킹 능력을 활용하면 더욱 좋은 기회를 만들 수도 있다고 생각한다. 어디까지나 작은 연구 분야에서 짧은 연구 경험과 다양한 경력 개발 활동에서 나온 제한된 경험이기에 실제 현장과 차이가 있다는 것을 이해해 주기를 바라면서, 재독과학기술인협회의 발전을 기원한다.

References

1. 이순칠, 퀀텀의세계(2022)
2. 양자전략위원회, 양자이니셔티브 (2025)
3. 홍형득, 과학기술정책론 (2016)
4. KIST, 21세기 Center of Excellence(2006)

한계를 넘는 엔지니어의 여정 (부제: 연구실, 군용시스템, 그리고 양자컴퓨터까지)

Jae-Ihn Kim[1]

[1]TopticaPhotonics SE, Lochhamer Schlag 19, 82166 Gräfelfing, Deutschland

sinerkim@gmail.com

Abstract. 독일에서 원자물리로 학위를 받은 후 군수 및 양자기술 산업계에서 다양하게 습득한 경험들을 공유하고자 한다. 이를 통해서 재독과협 회원들이 산업쪽으로의 진로에 대해 고민하고 결정이 필요할 때 참고사항으로 사용되길 바란다.

Keywords: 산업계진로, 군용시스템, 양자기술생태계

서론

진로 고민은 평생 하는 것 같다. 대학졸업 후 취업과 진학의 선택, 진학을 선택한 경우 졸업 후 학계에 남을 것인지 산업계로 가야 할지 결정해야 한다. 산업계로 가면 고민은 끝나는 걸까? 아니다. 이후 직장에서 치열하게 생존경쟁을 벌이며 본인의 경쟁력을 높이고 끊임없이 자신에게 본인의 일이 자신의 가슴을 설레게 하는지 항상 질문을 던지며 현직장에 머물든지 아니면 다른 곳으로 이직을 선택한다.

본인은 원자물리로 학위를 받고, 군용레이저 개발로 8년, 원자 기반 양자컴퓨터를 위한 광주파수 빗 개발로 5년이상, 총 13년 이상의 기간동안 산업계에서 경력을 쌓아왔다. 본 기고문을 통해 산업계로의 진로결정 시 고려했던 점들을 공유함으로써 재독과협 회원들이 산업 쪽 진로 결정할 때 작은 참고사항이 되었으면 좋겠다.

본론

군수산업으로의 결정 (독일에서 한국으로)

군용 레이저 개발이라는 분야는 나에게는 너무 낯설지만 굉장히 흥미로운 분야였다. 학위 과정에서 원자의 움직임이나 양자상태를 제어하는데 수 많은 레이저가 필요했고, 본격적인 실험 데이터를 얻기 위해서 레이저들의 상태를 확인하고 레이저 주파수를 원자 샘플에 맞추 잠금시켜야 했는데, 이 과정이 복잡해서 아침부터 시작해서 오후 늦게나 되어서야 끝날 수 있었다. 그리고 이를 이변이 없는 한 매일 해야 했다. 하지만 군용 레이저는 재정렬 과정이 전혀 필요 없어야 했으며 게다가 군 운용에 적합한 기준을 만족시켜야 했는데, 이것이 MIL-STD-810 같은 군용규격이다. 이에 따르면 고온, 저온 저장/운용, 충격, 진동 조건에서 시스템 또는 모듈이 문제없이 작동되어야 한다. 예를 들어 고온(저온) 저장의 경우 1주일 이상의 기간동안 70도(-40도)의 온도에 저장한 후 작동 시켰을 때 문제가 없어야 한다. 충격의 경우 바닥에 세게 집어 던지는 충격에도 레이저가 성능충족을 해야 하며, 바로 옆에 큰 기관총이 발사가 되는 진동에도 레이저가 문제없이 작동되어야 한다. 당시 대학 실험실에서의 경험만 있었던 나로서는 이런 조건은 한번도 상상해보지 못한 부분이었다. 이런 부분에 이끌려 (주)한화/방산 종합연구소에 입사를 하게 되었고 그곳에서 군용레이저 시스템 개발에 몰두하였다. 소형무장헬기용 레이저 표적지시기 및 거리측정기 체계개발, 화학작용제 검출용 라만 분광기를 위한 단자외선 고체레이저 개발, 수중물체 탐지를 위한 녹

색 파장 라이다 개발, 레이저 대공무기 과제참여 등 남들이 잘 하지 않는 군용 레이저 시스템 개발을 경험할 수 있었다. 이 첫번째 결정의 주안점은 내가 선택한 그 분야가 나의 가슴을 뛰게 하는지 여부였고, 그러한 결정으로 독일에서 한국으로 가족과 함께 이주하게 되었다. 그리고 당시로서는 한국에 계속해서 있을 예정이었고 독일로 다시 나오는 것은 생각지도 않았다.

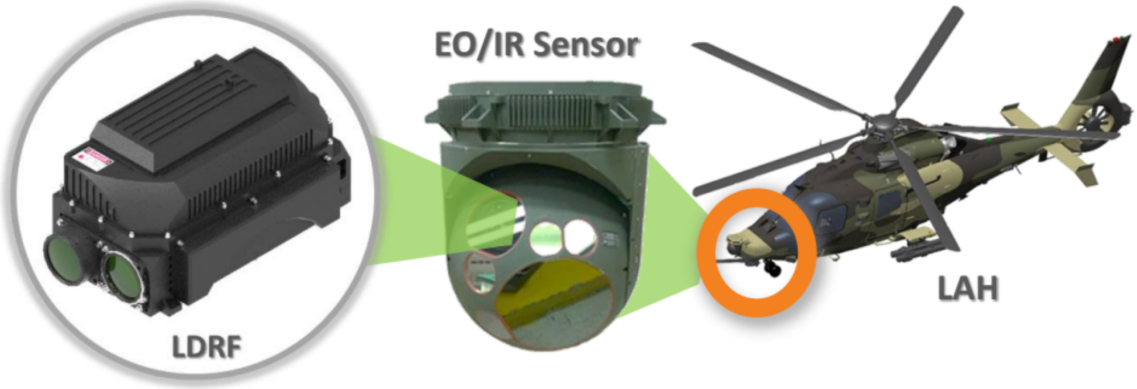


그림 1. 소형무장헬기용 레이저 표적지시기 및 거리측정기 체계개발

양자기술 분야로 결정 (한국에서 다시 독일로)

중간 관리자인 파트장으로써 거의 8년을 군용 레이저 개발에 몰두했다. 그쯤 되어서 나에게는 새로운 도전이 필요했으며, 첫 번째 결정을 했을 때와 달리 직업적인 고려사항뿐만 아니라 가족 차원에서의 고민도 더해졌다. 새로운 도전을 위해 독일 몇몇 회사와 연락이 되어서 면접도 보고 실패도 맛보고 연봉협상에서 결렬이 되었다. 그러다가 양자컴퓨터 및 양자센서 분야 생태계에서 독보적인 역할을 담당하는 Toptica Photonics에서의 인터뷰 기회를 얻게 되었다. 그리고 아내와도 독일 재이주에 합의가 되었지만 아이의 교육에 대한 고민도 생기게 되었다. 우리가 생각하고 매체에서 다루듯이 독일교육은 정말 좋은 것인가에 대해 질문도 많았었다. 결국엔 인터뷰도 성공적으로 마치고 최종 인터뷰를 위해 코로나 시기에 독일로 초청되어 갔을 때 참 많은 기분이 교차했다. 새 회사에서 맡은 업무는 optical frequency comb 개발이었다. 광주파수 빗에 대해서는 강의에서나 들어 봤었지만 개발업무를 시작하게 되어서 매우 도전적이었고 나 자신의 능력을 새롭게 증명해 보이는 계기를 잡게 된 것 같아 흥분되었던 것 같다. 광주파수 빗은 라디오 주파수 또는 다른 안정화된 광주파수와 임의의 광주파수 간의 사이를 연결해주는 장치로서 라디오 주파수의 안정성 또는 안정화된 광주파수의 안정성을 임의의 광주파수 대역으로 옮겨주는 중요한 역할을 한다 [1]. 그림 2에서와 같이 Toptica Photonics에서는 광주파수 빗이 여러 다른 단일모드 레이저와 19인치 랙에 직접화 되어서 양자컴퓨터에 필요한 초 저잡음 레이저 솔루션을 제공하고 있다 [2]. 광주파수 빗 개발업무는 광학적인 부분 뿐만 아니라 RF 주파수와 광주파수를 연결하고 있기 때문에, 라디오 주파수 관련 전자회로 부분, CW 레이저 안정화, 종합적인 최종 성능평가를 담당해야 해서 나 자신의 한계를 시험해야 했고 지금까지 신뢰성 부분을 포함하여 개발업무를 잘 이어오고 있다고 판단된다 [3]. 두번째 선택에서의 주안점은 선택한 분야의 적합성 뿐만 아니라 가족과의 원활한 합의였다.

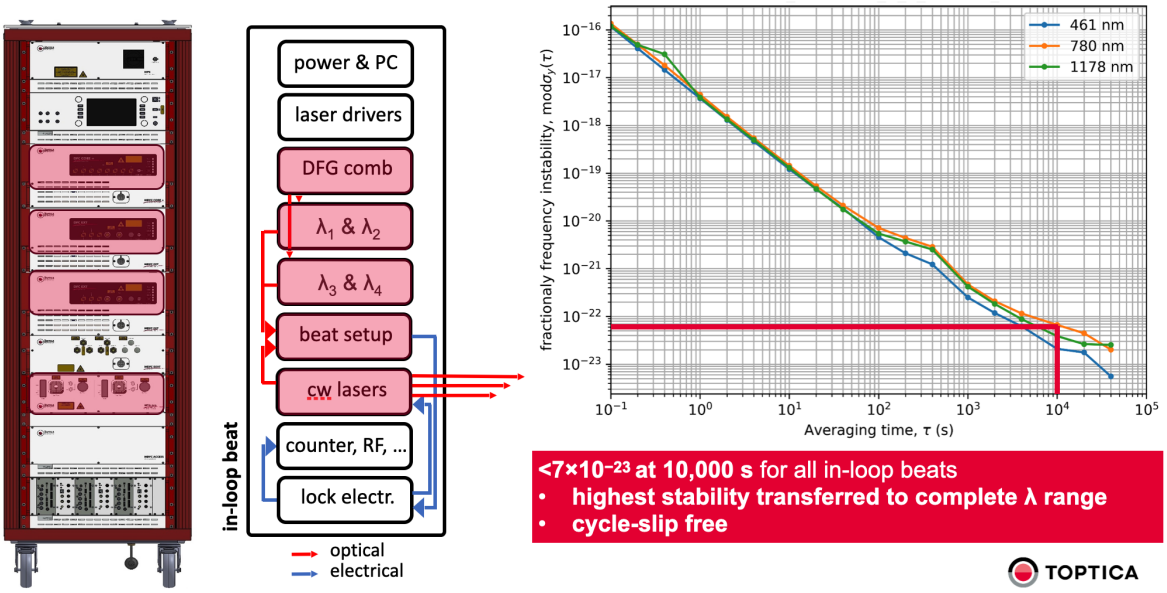


그림 2. 19인치 랙에 직접화된 광주파수 빗으로 안정화된 멀티컬러 레이저 시스템 [2]

독일회사와의 면접에서 몇 가지 느낀 점

위에서 언급했듯이 독일 회사와의 면접 과정에서 몇 가지 느낀 점을 공유하고자 한다. 일단 본인의 전문성이 확보되었고 서류 통과 후 면접을 보는 것으로 가정해보자. 기술면접에서는 서류에 나와 있는 전문성을 확인한다. 이는 본인의 일에 대해 면접관에게 설명하는 과정에서 검증을 받게 되는데, 예를 들어 가장 최근 중요하게 수행했던 과제에 대해 설명한다거나, 지원한 회사에서 수행할 주제에 대해 토의를 하는 과정을 통해 검증한다거나, 또는 전공 기초지식을 질문하는 경우도 있다. 그리고 그 회사의 대표적인 제품에 대해 설명해 보라는 질문도 있었다. 이러한 실무면접은 대개 화상면접으로 보는 경우가 많다. 따라서 화상회의가 원활하게 진행 될 수 있도록 기술적인 부분들이 잘 준비가 되어야 한다. 전공분야에 대한 전문성이 확인이 되면, 다음으로는 지원한 회사 팀원들과 원활히 지낼 수 있는지 여부를 on-site 면접을 통해 확인한다. 사실 이 부분이 중요한 역할을 하는데 재직중인 회사 내에서 면접과정을 지켜보면 이 부분에 최종합격 여부가 갈리는 경우가 많다. 소위 티키타카가 잘 되는 사람이 대부분 최종선택을 받고, 이 부분들을 최종결정자가 면접에 참여했던 모든 팀원에게 확인을 받는다.

결론

학위 후의 본인이 산업계에서 어떻게 캐리어를 성장시켰는지에 대해 기술하였다. 진로결정을 할 때 어떤 점에 주안점을 두었고 어떤 점을 고려하였는지 제시하였다. 또한 독일 회사에서 진행되었던 면접의 과정에서 느낀점을 기술하였다. 개인의 이런 경험들이 현재 산업 쪽 진로를 고민하는 젊은 학생들이나, 늦은 나이에 이직을 고민하는 재독과협 회원들에게 작은 도움이 되었으면 좋겠다. 나의 다음 캐리어는 어디가 될 것인지 현재도 끊임 없이 고민하고 준비 중에 있는데 기대된다. 혹시 또 다시 한국으로 갈 것인가?

References

1. S. A. Diddams, K. Vahala, and T. Udem, “Optical frequency combs: Coherently uniting the electromagnetic spectrum”, Science, **369**, 267 (2020)

2. www.toptica.com

3. J. Kim *et al.*, “Sub-kHz, RF-traceable frequency comb with dual-domain stabilization and automated locking”, EFTF April 23, Noordwijk, Netherlands (2026)

대학교와 대학원의 차이 (부제: 학사/석사/박사 학위란?)

Junbeom Park[1]

[1]Institute of Energy Technologies - Fundamental Electrochemistry (IET-1),
Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52425 Jülich, Germany
j.park@fz-juelich.de

Abstract. 우리는 대학원 과정을 “깊은 공부를 하고 연구를 한다” 같은 짧은 표현으로 인식하곤 한다.본 글에서는 학부과정과 대학원과정을 비교하며, 학부가 검증된 지식을 배우며 기초를 다지는 과정이고, 대학원은 기존 지식을 확장하여 '나만의 새로운 지식'을 만들어가는 과정의 의미를 전하고자 한다. 대학원 과정에서 마주칠 연구 조사부터 논문 작성까지의 고단한 과정은 결코 쉽지 않지만, 이를 통해 길러진 문제 해결 능력은 향후 어떤 도전 앞에서도 삶의 든든한 무기가 되어줄 것이다.

Keywords: 대학교, 대학원, 학사, 석사, 박사, 학위

1. 도입

2022년도 OECD자료[1]에 따르면 한국의 대학교 진학률은 약 70% 인데, 이는 OECD 회원국 중 최고수치이며, 평균인 약 47 %의 1.5배 수준의 높은 비율이다. 교육통계연구센터의 교육통계서비스[2]를 통해 지난 20여년간 한국의 진학률 및 학위과정별 졸업자 수를 정리해 보았는데, 그 기간동안 한국의 대학진학률은 약 65 - 85 %를 기록하며, OECD 회원국의 평균수치에 비해서는 지속적으로 높았음을 알 수 있다. 대한민국 국민의 4명중 3명이 대학을 진학하는 상황에서, 학사/석사/박사학위 졸업생 수의 변화도 같이 짚어보았다. 학사과정은 약 2015년부터, 석사과정은 약 2010년부터 졸업생 수가 더이상 늘지 않는데, 박사과정의 경우는 지난 20여년간 꾸준히 증가해왔음을 알 수 있다.

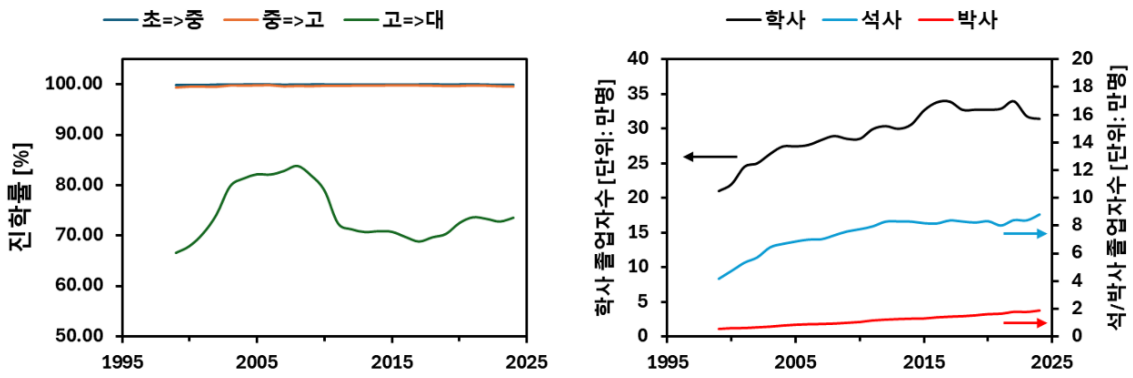


그림 1. 한국의 상위학교 진학률 및 학위과정별 졸업자수 [1,2]

통계를 보면 대학원 진학자 (특히 박사과정)의 숫자는 늘어나는데, 과연 대학원의 의미를 제대로 알고 진학하는 사람은 몇이나 될까? 저자가 대학교에 다닐때 들었던 얘기는, 대학교때 공부를 잘하던 사람은 대학원에 진학하는게 좋고, 공부를 못하던 사람은 대학원을 포기하고 회사등 다른 진로를 찾으라는 얘기였다. 솔직하게 얘기하면 저자는 공부를 못하는 편이였으나, 여러 이유로 대학원을 진학했고, 지금까지도 학계에 남아있다. 그동안 저자는후배들을 지도하면서, 학부과정과 대학원과정이 이런게 다르구나를 느끼고, 아래와 같이개인적인 생각을 정리해보았다.

2. 학부과정과 대학원 과정의 차이

2.1 학부과정

우리는 학부과정동안 해당 학과의 다양한 수업들을 듣고 경험하면서, 그 학과의 필수 지식을 쌓고 그중 일부 분야에 대해 조금 더 깊은 지식을 쌓게 된다. 이 과정에서 교과서를 비롯해, 백과사전, 논문, 실습서 등등 다양한 문헌종류를 습득하게 되고, 각각에서 내가 원하는 정보를 어떻게 찾고 공부하는지 등을 경험하게 된다. 방법적인 측면에서는, 고등학교때까지 우리가 열심히 해온 것과 크게 다르지 않다. 단지, 다루는 내용의 깊이가 있거나 진도가 나가는 속도가 빨라서 상대적으로 어렵다라고 느끼는 것이다. 가장 어렵게 느껴지는 때는 당연히 시험을 보는 시점인데, 저자의 대학생활때 자연계열은 이해/암기에 방점이 찍혀있어서 높은 난이도의 시험문제들이 출제(ex. 1주일간 동료들과 토론을 하면서까지 풀어내라는 수학과 시험문제)되는데, 공학계열은 실무능력에 방점이 찍혀있어서 전자기기를 제외한 자료를 시험장에 들고오는 오픈북 시험(ex. 선배들이 모아준 족보들도 인쇄물이므로 지참이 가능함)이 주로 진행되었다. 세세한 부분은 차이가 있을 수 있지만, 결국 존재하는 지식에 대해 문제가 나오고, 답을 하는 방식은 고등학교때까지 경험한 것과 다르지 않다.

일반적으로, 한국 기준으로는 4년제 대학에서 정해진 교육과정을 모두 이수하면 **학사학위 (Bachelor's degree)**를 받게 되고, 2/3년제 전문대학을 졸업하면 전문학사/준학사 학위를 받는다. 한국에서도 전공에 따라서 기간이 달라지는데 (ex. 건축학과 5년과정, 의과대학 6년과정 등), 해외 대학들의 경우 전공이나 지역에 따라 3년에서 6년까지 기간이 다양하다.

2.2 대학원과정

대학원에 입학하면서 보통 어떤 교수님/박사님 연구실에 소속이 되고, 해당 연구실에서 수행하는 연구와 연결된 연구주제를 받게된다. 다만, 한국 대학원생들은 대학원 수업을 일부 학점만큼 필수로 수강(석/박사 과정 모두)하고, 남은 시간동안 연구활동을 진행하게 된다. 해외 대학원의 경우 (보통 유럽지역) 박사과정때는 대학원수업이 없는 경우도 존재한다. 대학원 수업 수강은, 방법적으로는 학부과정과 동일하므로 여기서는 서술하지 않겠다. 남은 연구활동에 대해서 설명하면 크게 **1) 문헌검색 및 조사** (주로 학술논문들), **2) 연구 수행** (실험이나 시뮬레이션 등), **3) 문서 작성** (학술논문, 보고서, 과제제안서 등) 등으로 구분될 수 있다.

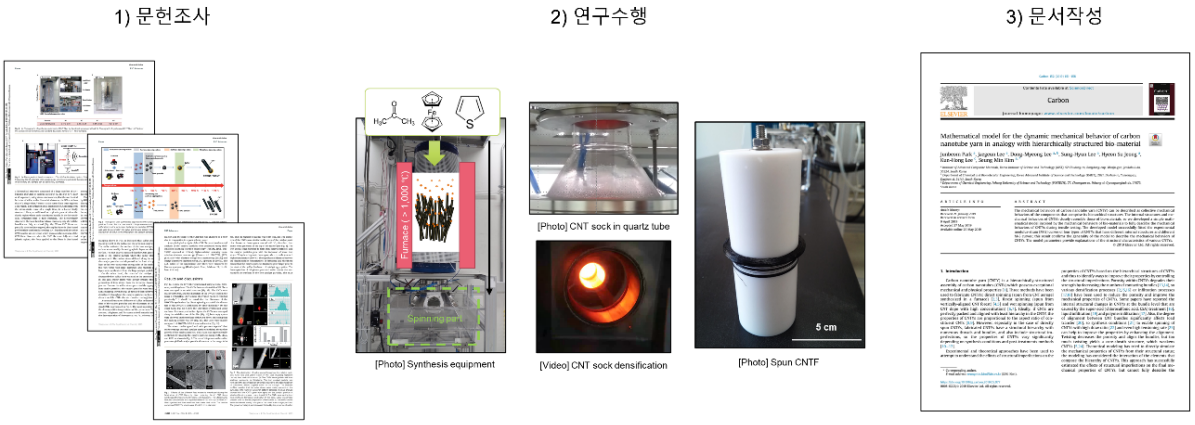


그림 2. 저자의 박사학위과정동안 수행한 연구활동들

1) 문헌검색 및 조사

본인의 연구주제에 대해, 기존의 연구자들이 어떤 연구를 진행했는지, 문제점은 무엇이였는지 등에 대해 알아가는 과정이다. 즉, 학부과정에서 익힌 '만들어진 지식의 습득'을 활용하는 과정이다. 다만, 학부때는 주로 정리가 잘 된 교과서의 지식을 습득하는 과정이였다면, 대학원에서는 상대적으로 빠르게 학계에 요약보고되는 학술논문들로부터 지식을 습득해야 한다. 방법은 동일하나, 학술논문의 목적자체가 그 분야의 연구자들끼리 빠르게 새로 얻은 지식을 공유하는 목적이므로, 그분야 사람들이 아니면 쉽게 이해하기 어렵다. 처음 대학원에 들어와서 논문 읽는 법 자체에 고생하는 사람들이 적지 않으나, 보통 학교 도서관에서 학술논문 읽는 법 등의 유용한 세미나가 많이 열리니 찾아서 잘 수강하길 바란다. 보통 학부과정에서 공부를 잘했던 학생들은 1)도 잘 하는 편이여서, 달리기 시합에서 출발선 앞쪽에서 시작한다고 표현하기도 한다.

2) 연구 수행

1)을 통해 기존의 연구들에서 빠진 부분이 무엇인지를 알았다면, 문제점을 해결할 방법을 생각해내고 이를 뒷받침할 증거들(실험결과, 시뮬레이션 결과 등)을 확보하는 과정이다. 연구실에서 기존에 자주 사용해왔던 도구들이 있으면, 도구의 사용법만 잘 익히면 증거확보 자체는 상대적으로 쉬운 편이다. 그러나 대학원 과정이 새로운 지식을 만드는 과정인 만큼 쉬운 것들은 보통 내가 발전시킨 것이 아니여서, 결국 쉬운 도구들을 이용해 무언가는 어려운 것을 해내야한다. 반대로 연구실에서 새롭게 도전하는 주제의 경우에는, 어떤 도구가 필요한지부터 찾아내고 설치하는 등의 어려움이 있지만, 문제를 해결해가는 과정에서 배운 노하우들은 언젠가 도움이 된다. 여기부터는 나만의 아이디어를 만들어내는 (사실 쥐어짜내는) 과정의 시작이다. 그렇기에, 앞에서 비유한 달리기 시합을 가져오면, 시합 중 발생하는 수많은 상황들로 인해 언제든지 추월하기도 뒤쳐지기도 하게 된다. 사실 그렇기 때문에, 남과의 비교는 무의미해지고, 나와의 싸움이 시작된다고 보는 것이 옳다.

3) 문서 작성

2)를 통해 알게된 새로운 내용들을 문서로 작성해서 다른 연구자들에게 알리며, 그 분야에 기여를 하는 과정이다. 대학원생의 경우 학술논문작성이 많은 비중을 차지하지만, 경우에 따라서는 연구수행에 필요한 금전적 지원을 해준 기관에 연구결과를 보고하는 문서를 작성하기도 하고, 새로운 연구비를 확보하기 위해 연구과제를 기획하는 문서를 작성하기도 한다. 또한, 학술논문과 별개로 특허를 작성하기도 한다. 각 문서마다 목적과 형식이 다르므로, 같은 연구결과를 가지고도 앞에서 말한 서로 다른 4가지 문서를 작성하게 되는 경우도 발생한다. 2)에서 얻은 나만의 아이디어나 지식을 글로 잘 표현해야 하는데, 여차하면 '나만 이해하는 글'로 빠져버리기 쉬워서, 남들이 이해할 수 있는 속도와 방법으로 글을 작성해야 한다. 1)과 마찬가지로 글쓰기 강좌는 학교도서관에서 자주 세미나가 열리니 꼭 찾아서 수강하길 바란다.

2.3 학위논문 발표 (디펜스)

앞에서 말한 1)에서 3)까지의 연구활동을 하면서 나만의 지식을 충분히 쌓았다라고 하면, 그 뒤 최종적으로 학위논문 발표(디펜스) 및 학위논문 제출을 하면서 **석사학위 (Master's degree)** 또는 **박사학위 (PhD, Doctor of Philosophy)**를 받게 된다. 참고로, 학위논문 발표/제출이 없이 졸업을 하는 상황이 발생하는데, 이를 석사수료 또는 박사수료로 부른다. 학위심사기준은 크게 보면 '나만의 지식'의 깊이와 과학적인 논리근거, 그 분야에 대한 기여도 등이 있다. 보통 **석사학위 (Master's degree)** 취득까지는 2년이 소요되고 늦어도 3년이면 학위를 받게 된다. 석사학위는 그 분야의 충분한 지식을 지니고 있다라는 뜻이 강해서, 성실히만 연구책임자의 지도를 잘 따라왔다면 큰 문제없이 학위를 받게 될 것이다. 문제는 **박사학위 (PhD, Doctor of Philosophy)**이다. 보통 4년과정이긴 하나, 기약이 없다. 박사학위의 영문명에 들어있는 Philosophy '철학'으로 생각해보면, 해당 분야에서 자신만의 깊이있는 논리로 새로운 지식을 창출해야하는 능력을 지닌 사람만이 박사학위 취득이 가능한 것이다. 개인적인 기준점으로는 학위과정 지도자에게 나만의 지식을 논리적으로 설득할 수 있고, 연구주제에 대해 자유롭게 토의를 할 수 있다면 박사학위를 받을 자격이 있다고 생각한다.

3. 결론

'아는 만큼 보인다'라는 표현이 있다. 나는 이 표현을 '알아가는 과정'과 '보는 과정'으로나누어 보면, 학부과정과 대학원과정의 차이를 잘 나타낼 수 있다고 생각한다. 학부과정에서는 '알아가는 과정'이 기존 지식을 공부하는 과정이고, '보는 과정'이 시험을 통해 검증하는 과정이라고 얘기할 수 있다. 대학원과정에서는 '알아가는 과정'이 기존 지식을 공부하는 것 뿐만 아니라 그 지식들과 내 연구활동 경험 모두를 엮어가며 확장하는 것이고, '보는 과정'이 논리정연하게 과학적으로 의미있는 나만의 지식을 만들어가는 과정이라고 얘기할 수 있다. 그래서 개인적으로 생각하기를, '나만의 지식'을 만들어가는 과정을 제대로 경험해본 사람은, 얼마든지 다른 새로운 분야에서도 핵심을 파악하고 도전을 해서 문제를 해결해 낼 수 있다고 믿는다. 물론 그 과정이 절대 쉬운 길은 아니지만, 도전해 볼 가치는 충분하다고 생각한다. **어차피 인생이라는 것도 계속 나에게 해결하기 어려운 문제를 던지는 과정이니까 말이다.**

4. 저자의 말

본 글은 저자의 개인 블로그에 올려두었던 글[3]을 거의 편집없이 옮겼음을 밝힌다. 개인 블로그에 잠들어 있던 글이, 특별논문집을 통해 빛을 보게 되었다고 생각해서 감사한 마음이다. 이 글을 읽는 차세대 과학자들에게 도움이 될만한 저자의 다른 글[4-6]도 전하며 여러분의 미래에 도움이 되길 바란다.

References

1. OECD 대학교 진학률 자료, https://www.oecd.org/en/data/indicators/population-with-tertiary-education.html
2. 교육통계서비스, 교육통계연구센터, https://kess.kedi.re.kr/index
3. 박준범, 대학교와 대학원의 차이 (부제: 학사/석사/박사 학위란?) https://junbeom-park-fzj.github.io
4. 박준범, 나의 유럽진출 경험담 인터뷰, https://k-erc.eu/bbs/board.php?bo_table=supporters&wr_id=68
5. 박준범, (독일) 율리히연구소 (Forschungszentrum Jülich, FZJ) 소개, https://k-erc.eu/bbs/board.php?bo_table=supporters&wr_id=62
6. 박준범등, 유럽주요국가별인재양성지원제도 (독일편), https://k-erc.eu/bbs/board.php?bo_table=publications&wr_id=52

계획대로 되지 않아도 길은 이어진다.

박형원

중앙대학교첨단소재공학과
hpark84@cau.ac.kr

1. 시작하는 말

연구자의 경력은 얼마나 치밀하게 계획할 수 있을까. 지금까지 걸어온 길을 돌아보면, 나의 경력은 계획을 차례로 실현한 결과라기보다 예상하지 못한 선택이 이어진 결과에 가깝다. 첫 미국 유학 지원에서 모두 탈락했고, 재도전 끝에 풀브라이트 장학생으로 MIT에 진학했다. 미국에서 박사과정을 계속할 계획이었지만 독일로 향했고, 예상보다 긴 박사과정을 거쳐 Fraunhofer IKTS에서 산업 응용 연구를 경험했다. 이후 여러 차례의 도전 끝에 중앙대학교 교수로 임용되어 한국으로 돌아왔다.

계획은 여러 번 어긋났지만 한 가지 목표는 변하지 않았다. 연구자의 길을 걷기 시작한 때부터 언젠가 독립적인 연구자로서 후학을 양성하고 싶었다. 그 목표에 이르는 과정에서 내가 지키려고 했던 원칙은 단순했다. 계획대로 되지 않더라도 포기하지 않고, 주어진 상황에서 선택할 수 있는 최선의 길을 고른 뒤 그 선택에 최선을 다하는 것이었다.

2. 유학은 목적이 아니라 수단이어야 한다

연구자를 꿈꾸기 시작하면서 해외 유학을 자연스러운 과정으로 생각했다. 학부연구생 시절 국내외 논문을 읽으며 연구를 바라보는 관점의 차이를 느꼈고, 연구적 식견뿐 아니라 문화적인 시야도 넓히고 싶었다. 그러나 석사과정 마지막 학기에 지원한 미국 대학에서 모두 불합격 통보를 받았다. 이후 약 1년 반 동안 재도전을 준비하면서 외부 장학금이 입학 가능성을 높이는 데 도움이 될 수 있다고 판단해 풀브라이트 장학금에 지원했다. 첫 지원에서 사용했던 연구계획서와 자기소개서를 다시 분석하고 보완한 결과 장학생으로 선발될 수 있었다. 당시의 실패는 아팠지만, 실패한 지원서를 다시 들여다본 경험이 다음 기회를 가능하게 했다.

MIT에 진학한 뒤에는 또 다른 현실과 마주했다. 한국에서는 나보다 뛰어난 사람과의 차이를 더 많은 시간과 노력으로 극복할 수 있다고 생각했다. 그러나 MIT에는 나보다 뛰어난 사람들이 나보다 더 치열하게 노력하고 있었다. 한번 벌어진 차이를 따라잡기 어려웠고, 그 과정에서 연구자로서의 자신감도 크게 흔들렸다.이 경험을 통해 유학 자체가 목표가 되어서는 안 된다는 사실을 배웠다. 나 역시 처음에는 '해외에 나가 학위를 받아야 한다'는 생각이 앞섰고, 그곳에서 어떤 연구를 통해 무엇을 이루고 싶은지는 충분히 고민하지 못했다. 유학을 준비하는 후배라면 어느 나라의 어느 대학에 갈 것인지를 고민하기 전에 다음 질문에 먼저 답할 수 있어야 한다.

'나는 왜 이 연구를 하려는가? 그곳에서 무엇을 배우고 어떤 연구자로 성장하고 싶은가?'

유명 대학에 입학했다는 사실만으로 연구자가 성장하는 것은 아니다. 낯선 문화 속에서 이방인으로 살아가며 연구 외적인 어려움까지 감당해야 한다. 새로운 문화에 들어가 사람들과 어울리려는 마음가짐과 외로움에도 쉽게 흔들리지 않는 내면의 힘이 필요하다.

3. 실패의 원인을 직시해야 한다.

MIT 졸업 후에는 미국에서 박사과정을 계속할 계획이었다. 실제로 몇몇 대학에서 입학 허가도 받았다. 그러나 그 무렵 결혼했고, 아내가 독일의 회사로부터 제안을 받으면서 계획이 달라졌다. 두 사람이 함께 일과 학업을 이어갈 수 있는 곳을 고민한 끝에 독일 FAU에서 박사과정을 시작했다.

독일에서 처음 맡은 연구는 conductive-bridging memory에 관한 것이었다. 새로운 물질에 대한 이해가 충분하지 않은 상태에서 '이렇게 하면 되겠지'라는 생각으로 연구를 시작했다. 약 20개월 동안 간헐적으로 희망적인 결과가 나타났기에 연구를 쉽게 중단하지 못했다. 그러나 최종적으로는 처음 선택한 물질의 특성상 원하는 메모리 성능을 구현하기 어렵다는 결론에 도달했다. 충분한 사전조사가 있었다면 피할 수 있는 실패였다. 이 경험은 모든 실패가 불가피하거나 낭만적인 것은 아니라는 사실을 가르쳐 주었다. 연구에서 시행착오는 피할 수 없지만, 기본적인 조사의 부족으로 발생한 실패까지 미화해서는 안 된다. 좋은 연구자는 실패를 많이 경험한 사람이 아니라 실패의 원인을 정확히 파악하고 같은 실패를 반복하지 않는 사람이라고 생각한다.

그렇다고 그 시간이 완전히 사라진 것은 아니었다. 20개월 동안 축적한 소자 제작과 측정 경험은 이후 연구의 밑바탕이 되었다. 실패가 자동으로 자산이 되는 것은 아니다. 실패의 원인을 정직하게 분석하고 그 과정에서 얻은 지식과 기술을 다음 연구에 연결할 때 비로소 경험이 된다. 성공적인 연구 성과도 중요한 전환점이 되었다. 초반의 약 2년을 실패한 프로젝트에 쏟은 뒤 새롭게 시작한 연구가 Nature Sustainability에 게재되었다. 이 성과는 연구자로서 소질을 의심하던 나에게 이 길이 나의 길이 맞다는 확신을 되찾아 주었다.

4. 전문성은 분야 사이의 연결에서 나온다.

나의 연구 분야는 광전자소자에서 시작해 표면화학과 계면공학으로 확장되었다. 겉으로는 큰 변화처럼 보이지만 두 분야는 단절되어 있지 않다. 광전자소자의 성능을 좌우하는 중요한 현상 역시 소재의 표면과 계면에서 일어난다. 소자를 연구하면서 표면화학에 대한 이해가 부족하다는 사실을 절감했고, 그 부족함이 오히려 새로운 분야에 도전하는 동력이 되었다. 그 결과 소자의 설계와 제작뿐 아니라 소재 표면에서 일어나는 화학적 상호작용까지 함께 이해할 수 있게 되었다. 기존 분야를 버리고 새로운 분야를 선택한 것이 아니라, 기존 연구에서 해결하지 못한 질문을 따라 전문성을 확장한 것이다.

선택적 가스센서 연구는 이러한 경험이 독립적인 연구 아이디어로 발전한 사례다. 가스센서 연구자들과 대화하던 중, 표면의 화학적 상호작용을 이용하면 특정 가스에 대한 선택성을 구현할 수 있겠다는 생각이 들었다. 표면화학에서 추상적으로 논의되던 상호작용을 센서의 측정 결과로 보여주고 싶었다. 아이디어의 고안부터 실험 설계와 논문화까지 주도하면서 비로소 나만의 연구 질문을 만들 수 있다는 자신감을 얻었다.

오늘날의 융합연구에서는 한 사람이 모든 분야에서 깊은 전문성을 갖추기 어렵다. 따라서 자신의 핵심 전문성을 갖추는 동시에 다른 분야의 언어를 이해하려는 노력이 필요하다. 실험 연구자와 계산과학자는 서로의 어려움을 충분히 이해하지 못할 수 있으며, 같은 용어도 분야에 따라 다르게 사용된다. 이러한 간극이 해소되지 않으면 상대방의 결과에 대한 불신으로 이어질 수 있다. 또한, 성공적인 공동연구를 위해서는 각 분야의 깊이 있는 연구자들을 연결하고 서로의 언

어를 통역하는 역량이 필요하다. 모든 분야를 깊이 알 수는 없더라도 상대방의 연구를 이해하고 자신의 분야와 연결할 수 있을 만큼 폭넓게 공부해야 한다. 장기적인 협력에서는 당장의 이익만을 앞세우기보다 때로는 양보할 줄도 알아야 한다. 연구협력 역시 결국 사람과 사람 사이의 관계이기 때문이다.

5. 학계와 산업계의 차이를 경험하다.

박사학위 취득 후 대학에서 박사후연구원(포닥)으로 근무하다가 새로운 연구환경을 경험하고자 Fraunhofer IKTS로 자리를 옮겼다. 그곳에서는 산업계와 가까운 연구가 학계의 연구와 어떻게 다른지 배울 수 있었다. 학계에서는 새로운 현상을 발견하면 논문을 통해 신속하게 공개하는 것을 중요한 성과로 여긴다. 반면 산업계와 연계된 연구에서는 경쟁력 있는 결과를 대외비로 보호하기도 한다. 학계가 지식을 공개하고 확장하는 데 무게를 둔다면, 산업계는 기술을 이용해 실질적인 가치를 창출해야 한다. 어느 한쪽이 더 우수한 것이 아니라 연구의 목적과 성과를 판단하는 기준이 다른 것이다. 이 경험은 내가 학계와 산업계 가운데 어느 쪽에 더 가까운 사람인지 판단하는 계기가 되었다.

Fraunhofer에서 Horizon Europe 과제에 참여하며 형성한 유럽의 연구 네트워크는 귀국 후에도 중요한 자산이 되고 있다. 유럽 여러 나라의 연구자들을 국내 연구자들과 연결하면서 새로운 협력 기반을 만들고 있다. 국제협력은 단순히 명함을 많이 주고받는 것에서 시작되지 않는다. 상대 분야를 이해하려는 태도, 약속한 역할을 책임지는 자세, 단기적인 이익보다 장기적인 신뢰를 중시하는 태도에서 시작된다.

6. 계획보다 중요한 것은 다시 방향을 잡는 힘이다.

돌이켜보면 한국과 미국, 독일의 연구문화를 모두 경험하고 유럽의 국제공동연구 과제에 참여한 경력은 임용 과정에서 차별화된 강점으로 작용했다. 반면 오랜 해외생활로 국내 연구 네트워크가 부족하다는 점은 귀국 후 새롭게 해결해야 할 과제가 되었다. 해외 경력은 그 자체로 완성된 경쟁력이 아니다. 해외에서 얻은 지식과 관계를 국내 연구환경에 맞게 연결해야 비로소 실질적인 자산이 된다.

지난 나의 경력에서 계획대로 이루어진 일은 많지 않다. 첫 유학 지원에서 모두 떨어졌고, 재도전 끝에 예상하지 못했던 풀브라이트 장학생이 되었다. 미국에서 박사학위를 받을 계획이었지만 독일로 향했고, 박사과정은 생각보다 오래 걸렸다. Fraunhofer에서 산업 응용 연구와 유럽의 국제공동연구를 경험하게 될 것도 예상하지 못했다. 그럼에도 처음 품었던 목표에 도달할 수 있었던 이유는 계획이 어긋날 때마다 그 상황에서 선택할 수 있는 최선의 길을 찾고, 그 선택에 다시 최선을 다했기 때문이라고 생각한다. 연구자의 경력에 필요한 것은 완벽한 계획보다 계획이 틀어졌을 때 다시 방향을 잡는 힘인지도 모른다.

연구를 수행할 때에도 기본으로 돌아가려고 노력한다. 지금까지 좋은 성과로 이어진 연구들은 반드시 화려한 최신 기술을 사용한 연구가 아니었다. 오히려 고전적인 화학적 상호작용을 제대로 이해하고 이를 새로운 문제에 적용한 연구가 의미 있는 결과로 이어졌다. 최신 연구 경향을 파악하는 것은 중요하지만, 유행하는 기술을 좇느라 현상을 지배하는 기본 원리를 놓쳐서는 안 된다. 무엇이 새롭게 보이는지를 묻기 전에 왜 그런 현상이 발생하는지를 물어야 한다.

스티븐 호킹은 “Science is not only a disciple of reason but, also, one of romance and passion”이라고 말했

다. 과학은 이성의 학문인 동시에 낭만과 열정의 학문이라는 뜻이다. 낭만과 열정만으로 좋은 연구자가 될 수는 없다. 그러나 그것이 없다면 반복되는 실패와 불확실성을 견디며 긴 연구자의 길을 걷기도 어렵다. 나 역시 연구에 대한 낭만과 열정이 없었다면 여기까지 오지 못했을 것이다. 길은 예상보다 훨씬 멀었고 여러 번 뜻하지 않은 방향으로 굽어졌다. 그러나 멀리 돌아온 끝에, 지금 이 자리에 설 수 있었던 것 같다.

후배 연구자들에게 나의 경험을 통해 전하고 싶은 말은 하나다. 계획한 길이 막혔다고 해서 목표까지 사라진 것은 아니다. 그 순간에 할 수 있는 최선의 선택을 하고, 선택한 길에서 다시 최선을 다한다면 예상과 달라진 길도 결국 자신만의 경력이 될 수 있다.

전임 교원에 대해서 박사과정 때 미리 알았으면 좋았을 것들

남세광

[1] 서울시립대 인공지능학과, 02504 서울특별시 동대문구 서울시립대로 163
s.nam@uos.ac.kr

Abstract. 많은 박사과정 학생들은 전임교원이 될 수 있을지에 대한 막연한 불안을 안고 학위과정을 보내지만, 그 불안의 근거를 구체적으로 따져보는 경우는 드물다. 본 글은 그러한 전환을 최근에 마친 저자가 박사과정과 박사후과정을 되돌아보며, 전임교원 임용에 대해 널리 퍼져 있는 몇 가지 믿음이 실제와 다르다는 점을 지적한다. 논문 실적 만으로 임용 결과가 결정되지는 않는다. 학과는 새로운 동료를 선택하는 조직이므로 세부 전공 분야의 적합성, 강의 가능 과목, 해외 네트워크를 논문 실적과 함께 고려하며, 지원자는 이에 맞추어 지원 전략을 세워야 한다. 완벽한 지원서가 준비될 때까지 기다리는 태도나 자신의 연구 분야에 전망이 없다고 단정하는 태도 역시 도움이 되지 않는다. 이어서 저자가 실제로 유용하다고 판단하는 것들, 즉 박사과정을 '성공적으로'가 아니라 최대한 빨리 마치는 것, 박사후과정을 독립적인 연구자가 되기 위한 훈련 기간으로 활용하는 것, 기회가 왔을 때 일찍 부임하여 교수 사회의 실제 업무량과 국내 연구과제 체계를 익히는 것, 그리고 Research talk을 통해 자신과 연구주제를 적극적으로 홍보하는 것을 정리한다. 본 글은 저자의 개인적 경험에 기반한 기록이지만, 국내 대학의 전임교원을 지망하는 후배 학도에게 실질적인 지침이 되기를 기대한다.

Keywords: 전임교원 임용, 학술 경력 개발, 박사후과정, 연구 홍보

1. 서론

모든 박사과정 학생들은 한번쯤은 전임교원이 되는 것을 꿈꾼다. 저자 본인도 예외는 아니었다. 이 때문에, 대부분의 박사과정 학생들은 '그 꿈을 이룰 수 있을까' 하는 막연한 두려움과 그로 인한 심리적 압박감을 가지고 긴 연구의 터널을 걸어 나간다. 저자 본인도 예외는 아니었다. 박사과정이 시작하는 시점부터 끝난 5년 후까지 단 한번도 자신의 실력과 실적이 전임교원의 꿈에 더 가까이 다가간다고 느껴본 적이 없었으며, 그러한 심리가 정신적 스트레스로 이어지는 경우가 있었다. 전임 교원이 되고 난, 지금의 시점에서 과거를 돌아봤을 때, 예전에 생각했던 전임교원에 대한 생각과 지금의 생각이 많이 다름을 느낀다.

본 글은 대한민국 대학교의 전임교원을 꿈꾸는 후배 학도를 위하여 필요한 지원자의 마음가짐에 대해서 정리하고, 새로운 동료를 선택하는 학과의 입장에서 신임교원에게 기대하는 바에 대해서 서술하고자 한다. 물론, 본 글에 나온 제언이 전임교원을 초빙하고자 하는 모든 학과의 입장을 대변한다고는 볼 수 없으나, 최소한 이 글을 읽고 있는 전임교원 희망자에게는 도움이 될 만한 글이 될 것이다.

2. 전임교원이 되기 위해서 필요 없는 마음가짐들

논문만이 전임교원을 결정하는 절대적인 기준이 아니다. 물론, 양질의 논문을 쓰는 것은 지원자를 평가하는데 있어서 매우 중요한 척도이지만, 미래의 동료를 뽑는 학과 입장에서 지원자의 역량이 얼마나 매력적인지 생각해 봐야 한다. 구체적으로, 학과에서 원하는 최종 임용자의 조건은 학벌, 세부 전공 분야, 해외 네트워크로 크게 세가지 일 것이다. 여기에서 학벌은 우리가 노력해서 변하지 않는 것이기 때문에 제외하고, 세부 전공 분야는 전임교원 채용 공고문의 모집분야에 잘 설명되어 있으므로, 지원자 본인이 그 분야에 적합한 지원자인지 스스로 판단해야 한다. 이런 경우, 대개는 모집분야에 맞는 강의 가능 과목까지 같이 고려하여 면접에 임해야 한다. 마

지막으로 가끔씩 학과에서 해외 네트워크를 중요시 하는 경우가 있는데, 이것은 대형 인력양성 사업 또는 국제공동연구 사업을 준비하거나 이미 수주한 상태인 경우이다. 따라서, 지원자는 해외에서 공동연구가 가능한 연구자들에 대한 강조를 면접에서 해야 한다. 요약하면, 논문 중심의 사고에서 벗어나 학과입장에서 내가 어떻게 보일지 생각을 하고 지원 전략을 마련해야 한다는 것이다. 이러한 지원전략의 전환은, 일부 연구중심 대학을 제외하고는 전임교원이 되는데 있어서 더 도움이 될 것이다.

두번째 잘못된 생각은 '완벽한 상태에서 지원'을 한다는 마음가짐이다. 지원서를 준비하는 시간과 노력이 많이 들어가기 때문에, 처음부터 완벽한 지원서를 쓸 수는 없다. 그리고, 학교별로, 학과별로, 시기별로 원하는 인재의 기준이 다르기 때문에 자기소개서와 연구계획서에 많은 시간을 할애해야 하며, 이는 다른 서류 준비가 미리 되어 있어야 함을 의미한다. 따라서, 준비가 되어 있지 않음에도 지원을 미리 해 보는 것이 중요하다. 미리 지원을 하면 추가적인 정보를 얻을 수 있는데, 그것은 지원자의 현재 역량으로 어느 정도의 상위권 학교에서 서류 통과가 되는지 알 수 있다는 점이다. 실제로 저자 본인은 박사과정 말년차부터 지원을 시작하여 1년이 넘는 시간 동안 꾸준하게 자기소개서와 연구계획서를 다듬어 나갔다. 지원 학교의 스펙트럼도 넓게 가져가고 연구소와 회사 등에도 다양하게 지원하여 저자 본인의 객관적인 실력 파악을 하려 노력했다.

세번째 잘못된 생각은 '내 연구 분야가 전망이 없다'와 같은 부정적인 마음가짐이다. 더하여 지원자 스스로의 역량을 낮게 평가하는 마음가짐 또한 좋지 않다. 이는 전임교원이 되는데 아무런 도움이 되지 않을 뿐더러, 전임 교원이 되지 않았을 때를 대비한 다른 진로 탐색을 방해하는 역할만 할 뿐이다. 예를 들어 2021년 경, 저자가 독일에서 햅틱스를 주제로 박사과정을 하고 있을 때, 이 연구 주제는 주변의 사람들이 크게 주목하지 않는 분야였으나, 시간이 지난 후, 2026년 현재는 피지컬 AI 키워드에서 없어서는 안될 인기 학문으로 부상하였다. 또한, 당시 전자 물리학으로 독일에 박사후 과정으로 온 친구는 끝내 전임교원이 되지 못하고 한국의 반도체 대기업에 취업했으나, 현재는 어느 누구보다 높은 연봉을 자랑하며 만족스러운 회사생활을 하고 있다. 이처럼 미래는 아무도 모르며, 현재의 기준으로 미래를 어둡게 판단하는 것 만큼 어리석은 행동은 없다.

3. 전임교원이 되기 위해서 필요한 마음가짐들

본 절에서는 대한민국의 전임교원이 되기 위해서 필요한 행동들에 대해서 설명을 하고자 한다.

가장 먼저 해야 할 일은 박사과정을 빨리 끝내는 것이다. 박사과정이라는 것은 긴 터널과 같아서 자칫하면 스스로의 정신적 건강을 읽기 쉽다. 지도교수에게 장기간 배움을 받는 일은 실제로 겪어보니 쉬운 일이 아니며, 연구 동료와의 갈등까지 겹친다면 성공적인 끝맺음을 할 수 없는 상황이 생긴다. 나는 박사과정이 굳이 '성공적'일 필요가 없는 과정이라 생각한다. 성공여부 보다는 최대한 빨리 과정을 마치는 것이 중요하다고 생각한다. 장기적으로 수행한 연구를 장문의 글로 마무리 하는 능력을 배웠다면 그것으로 충분하며, 다음 과정을 가기 위한 기반이 되는 역할에 불과하다고 생각하는 것이 오히려 마음이 편하다.

오히려 박사후과정이 중요하다고 생각한다. 박사후과정은 전임교원에게 요구되는 '독립적인 연구자'로서 후배들을 가르치는 능력을 배양하는데 좋은 기회이다. 박사 학위를 취득했기 때문에, 지도교수가 그만큼의 대우를 해주어 대화할 때 더 존중을 받는다는 느낌도 있으며, 박사과정을 제대로 수행한 경우, 새로운 연구실의 박사과정 학생들이 가진 부족함이 잘 보여 제대로 된 도움을 줄 수 있을 것이다. 예를 들어, 저자 본인의 경우, 독일에서 박사학위를 받고 영국에서 박사후과정을 했을 때, 진학전에는 영국의 native English speaker들이 쓴 논문에 대해서 non-native

speaker인 내가 어떻게 논문 지도를 할 수 있을지에 대한 스스로의 의문이 많이 들었다. 그러나, 정작 그들이 쓴 논문을 보니, 글의 논리가 맞지 않아 내가 도와줄 수 있는 부분이 많다는 것을 깨달았다. 이러한 경험은 추후에 내가 학생지도를 자신 있게 할 수 있는 기반이 되었다.

간혹 100% 마음에 들지는 않지만, 전임교원에 최종 합격한 경우가 발생하며, 선택을 해야 하는 상황이 온다. 그런 상황이라면,취업을 빨리하는 것이 좋다고 생각한다. 교수 사회는 직접 경험하지 않으면 느낄 수 없는 많은 다름이 있다. 가장 대표적인 것이 업무량이다. 겉으로 보기에는 여유 있으며 존경받는 직업으로 보이지만, 현실은 연구, 수업, 학과행정을 동시에 해야 하는 hard work직업이다. 따라서, 기회가 왔을 때, 최대한 빨리 이러한 차이에 대해서 느끼고 적응해야, 가정도 이루고 육아도 병행해야 하는 추후에 개인 삶과의 밸런스를 맞출 수 있다고 본다. 또한 일찍 취업함으로써 얻는 호봉 및 연금의 상승은 말할 필요도 없다.

취업을 일찍하면, 국내 연구과제에 대한 파악을 빨리 할 수 있다는 장점도 있다. 교수라는 직업에 잘 뿌리 내리기 위해서는 연구과제의 선정이 중요한데, 전임 교원이 되기 전에는 어떤 과제가 있는지 따로 파악하기가 힘들다. 차라리, 장기간에 걸쳐 여러 과제에 제안서를 써보고, 평가를 받아보면서 매해 반복되는 여러 연구과제들의 특징을 파악하고 수주하는 것이 좋다.

마지막으로, 전임교원이 되기 위해서 가장 필요한 것은 자신과 연구주제에 대해서 홍보하는 것이다. 특히 박사후과정 동안에는 한국에 들어갈 기회가 있을 때마다 관심있는 학과의 교수님이나 교수가 된 지인에게 부탁하여 Research talk에 대한 기회를 최대한 많이 만들어야 한다. 이러한 방식은 초청하는 교수님이 오히려 반긴다. 요청을 하는 사람 입장에서는 상대에게 부담이 될까 말을 못 꺼내지만, 초청하는 입장에서는 발표자의 연구도 알고, 발표 분야에 대한 최신 동향을 알 수 있게 되고, 국제 네트워크의 강화까지 되는 1석 3조를 마다할 이유가 없다. 많은 발표를 통해 한국방문 때 지출한 왕복 비행기 값 정도는 스스로 벌어서 가는 적극적인 사람이 되자.

4. 결론

본 글은 대한민국 대학교의 전임교원이 되는 과정을, 그 전환을 최근에 마친 사람의 관점에서 되돌아본 것이다. 세 가지를 강조하고 싶다. 첫째, 논문 실적만을 중심에 둔 지원 전략은 불완전하다. 학과는 함께 일할 동료를 선택하는 것이므로, 논문 실적과 더불어 세부 전공 분야의 적합성, 강의 가능 과목, 해외 네트워크를 함께 평가한다. 둘째, 완벽한 지원서를 기다리거나 자신의 분야와 역량을 낮게 평가하는 태도는 유용한 피드백을 얻는 시점만 늦출 뿐이다. 일찍, 그리고 폭넓게 지원해 보는 것 자체가 자신의 위치를 객관적으로 가늠하는 방법이다. 셋째, 박사과정은 목적지가 아니라 다음 단계를 위한 기반으로 이해하는 것이 좋다. 학과가 요구하는 연구 독립성, 제도에 대한 이해, 그리고 인지도를 가장 직접적으로 쌓아주는 것은 박사후과정, 이른 임용, 그리고 자신의 연구에 대한 꾸준한 홍보이다. 물론 위의 제언 중 어느 것도 임용을 보장하지 않으며, 하나의 글이 모든 학과의 기대를 대변할 수도 없다. 저자가 전하고 싶은 것은 관점의 전환이다. 즉, '내가 충분히 잘하고 있는가'만이 문제가 아니라, '나를 선택할 동료들에게 내가 어떻게 보이는가'도 함께 물어야 한다는 것이다. 그렇게 바라보면 박사과정에 따라붙는 불안의 상당 부분은 다룰 수 있는 문제가 되고, 임용에 이르기까지의 시간은 확신을 기다리는 기간이 아니라 근거를 쌓아가는 기간이 될 수 있다.