

멀티스케일 계산을 위한 제일원리 전산 데이터 기반 머신 러닝 포텐셜 개발

강 준 희^{*,**} · 한 병 찬^{**,†}

^{*}한국에너지기술연구원 플랫폼연구실, ^{**}연세대학교 화공생명공학과

Development of First-Principles Database Driven Machine Learning Potential for Multi-scale Simulations

Joonhee Kang^{*,**} and Byungchan Han^{**,†}

^{*}Platform Technology Laboratory, Korea Institute of Energy Research, Daejeon 34129, Republic of Korea

^{**}Department of Chemical and Biomolecular Engineering, Yonsei University, Seoul 03722, Republic of Korea

Abstract: 최근 가공할만한 성능의 슈퍼컴퓨터에 머신 러닝 기법을 연동한 인공 지능형 소재 정보학이 과학 기술 및 산업계에 새로운 연구개발 패러다임으로 급속히 확산되고 있다. 본 기고문에서는 이 기법의 성공에 핵심적 요소인 정확한 데이터베이스 구축을 위해 제일원리 전산을 적용하는 것과 이를 기반으로 소재를 구성하는 원소 간 인공 신경망 포텐셜을 만드는 방법을 소개하고자 한다. 이 연구 방법론은 나노 스케일 신소재 개발에 적용할 경우, 양자역학 수준의 정밀도로 순수 제일원리 전산 대비 100배 이상의 빠른 결과를 도출할 가능성이 있음을 예시한다. 이는 향후 다양한 산업계에 막대한 파급효과를 가져올 것으로 예상된다.

Keywords: machine learning, density functional theory, database, nanoparticle

1. 서 론

최근 학계 및 산업계 전반에 인공지능을 이용한 머신 러닝에 대한 관심이 그 어느 때보다 높다. 특히, 지난 2016년 3월 구글 딥마인드에서 개발한 알파고가 세계적인 바둑기사 이세돌 9단을 4대1로 완파한 사건은 인공지능의 수준이 두려움을 자아낼 정도로 우리 곁에 성큼 다가왔음을 체감하게 했다. 사실, 그러한 광범위한 빅데이터에 기반한 머신 러닝 기법은 예전부터 많이 이용되어온 것이 사실이다. 작게는 스팸 이메일 필터링으로부터, 크게는 아마존과 구글 같은 대형 기업의 영업에도 적용되어 왔다.

이제 머신 러닝 기법은 사회적인 조직 문화를 넘어 과학 기술계에도 적용되어 난제에 대한 돌파

구를 제시하고 있다. 예를 들면, 신약 개발에 소요되는 막대한 시간적 금전적 비용의 획기적 감축, 자동차의 자율 주행, 영화와 드라마 추천 알고리즘, 의학적 진단 보조 등에 활용되고 있다. 특히, 급속히 고성능화 되고 있는 슈퍼 컴퓨터와 함께 양자역학 기반 제일원리 전산 프로그램의 발달이 머신 러닝과 연동되어, 광범위한 나노 스케일 소재 후보군에 대한 고속 스크리닝 같은 새로운 연구개발 패러다임을 구축하고 있다. 예를 들어, 연료전지 전극 촉매로 활용되는 백금의 나노 입자 크기에 따른 전기화학 안정성 예측[1], 다양한 2차원 소재의 접합 구조에서의 수소 생성 촉매 성능 고속 전산 스크리닝[2], 질소가 첨가된 폴리렌 구조 생성 예측을 통한 새로운 촉매 설계[3]에 대한 연구가 있다. 과거, 양자역학 기반의 밀도 범함수 이론 계산(density functional theory, DFT)과 제일원리 분자동역학은 컴퓨팅 자원의 과다 사용으로

저자 (E-mail: bchan@yonsei.ac.kr)

수 nm의 모델링만 가능했으나, 이러한 난관이 점차 극복되고 있다.

2007년 Behler와 Parrinello가 제시한 머신 러닝 기반 포텐셜(Behler and Parrinello neural network potential: BPNNP)[4]은 고속 멀티스케일 계산을 위한 새로운 방향을 제공했다. 이 방법은 나노 입자의 구조에 대한 정보를 대칭 함수로 맵핑한 후 DFT로 계산된 에너지를 인공 신경망 방법으로 학습시켜 포텐셜 라이브러리를 구축하는 것으로 양자역학 수준의 정확도를 유지하는 동시에 전통적 DFT 대비 약 100배 이상 빠른 계산 속도를 확보할 수 있어, 산화물과 복잡한 유기물 분자 구조까지 예측하는데 성공했다[5-9]. 본 고에서는 BPNNP 방법론에 대해서 소개하고 이를 활용한 합금 나노 입자 구조 예측 연구 결과를 보여주고자 한다. 이 중 합금 구조 예측에서 더 나아가 삼중 합금까지 안정한 원자 배열을 몬테카를로(Monte Carlo) 계산을 통해 예측하고 이 결과를 원자 크기 관점에서 분석했으며 표면의 결합 길이 변형률에 대해서 설명했다. 머신 러닝 기반 포텐셜은 기존 양자 계산과 고전 역학 바탕의 계산의 단점들을 극복하고 복잡한 구조 또는 큰 스케일의 계산에 적용될 것이며 새로운 공업화학 소재 설계에 활용될 것이다.

2. 머신 러닝 포텐셜 방법론

일반적으로, 단일 또는 다성분계 원자들로 구성된 시스템에 제일원리 DFT 계산을 적용하면 그 에너지와 안정한 구조를 예측할 수 있다. BPNNP 방법은 개별 원자의 에너지를 머신 러닝으로 예측하여 모든 구성 원자의 에너지 합이 DFT 예측 값과 같게 최적화 함으로써 그 포텐셜을 구축한다. 여기서 핵심 과정은 주어진 원자의 에너지는 주변의 일정거리(cut off radius) 이내의 원자들에 의해서만 결정된다는 것이다. 이를 수학적으로 컷오프(cut off) 함수로 만들어서 식 (1)과 같이 표현한다.

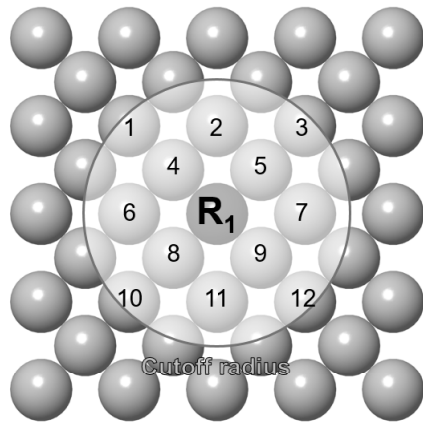


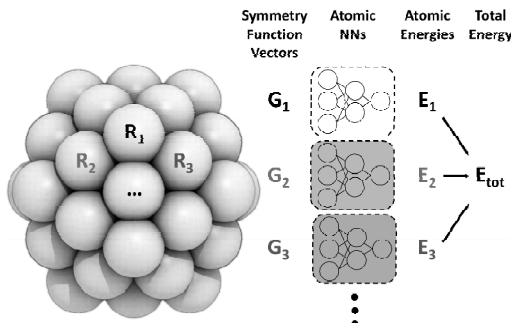
Figure 1. 기준원자 (R_1)의 컷오프 반지름 이내의 해당되는 원자를 표시한 예시.

$$f_c(R_{ij}) = \begin{cases} 0.5 \cdot \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) - 1 \right] & \text{for } R_{ij} \leq R_c \\ 0 & \text{for } R_{ij} > R_c \end{cases} \quad \text{식 (1)}$$

R_{ij} 는 기준 원자 (i)와 다른 원자 (j) 사이의 거리를 나타내고 R_c 는 컷오프 반지름을 나타낸다 (Figure 1). 컷오프 반지름을 벗어나는 곳에 있는 원자들은 에너지에 영향을 주지 않기 때문에 R_{ij} 가 R_c 보다 클 때는 함수 값이 0이 된다. 식 (1)과 같은 컷오프 함수를 바탕으로 주변 원자들 간의 영향을 표현하기 위해서 대칭 함수를 도입하게 된다. 대칭 함수를 고려하는 이유는 모든 원자들을 포함한 시스템 자체를 회전을 시켜도 각 원자들 간의 결합 길이나 결합 각도는 변화하지 않을 것이고 시스템의 에너지도 변화하지 않기 때문이다. 즉, 주어진 거리 범주 내 에너지를 원자간 결합 길이와 각도의 함수로 전개하는 것이다. 이를 표현하기 위해 먼저 식 (2)와 같이 결합 길이와 관련된 대칭 함수를 사용하게 된다.

$$G_i^2 = \sum_j e^{-\eta(R_{ij}-R_s)^2} \cdot f_c(R_{ij}) \quad \text{식 (2)}$$

길이 관련 대칭 함수는 가우스 형태의 분포함수로서, η 는 분포의 너비를 결정하는 변수이고 R_s 는



*출처: *Phys. Chem. Chem. Phys.*, **20**, 24539-24544 (2018).

Figure 2. 각 원자들 (R)의 대칭 함수 값 (G)을 구하고 인공 신경망을 통해 에너지를 구하는 방법 모식도.

분포의 중심을 결정하는 변수이다. 컷오프 함수를 같이 고려하기 때문에 중심 원자로부터 거리가 가까울수록 큰 영향을 주게 된다. 추가적으로 결합 각도를 고려하기 위해 식 (3)과 같이 결합 각도와 관련된 대칭 함수를 사용하게 된다.

$$G_i^{\lambda} = 2^{1-\zeta} \sum_{j,k \neq i}^{all} \frac{(1 + \lambda \cos \theta_{ijk})^{\zeta} \cdot e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)}}{f_c(R_{ij}) \cdot f_c(R_{ik})} \quad \text{식 (3)}$$

여기서 ζ 는 각도의 영향력을 결정(angular resolution)하고 λ 는 +1 또는 -1의 값을 가지며 코사인 함수의 0°와 180°일 때 최대값을 변환시키는 역할을 한다. 역시 컷오프 함수를 같이 고려하며 가까운 원자들의 영향이 더 크게 된다. 입력 변수들을 추가하면, 기준 원자로부터 주변 환경을 더 정밀하게 모사할 수 있으며, 대칭 함수의 값이 아주 비슷할 경우 구조상 차이가 미미하다는 것을 의미한다. 이렇게, 주어진 원자 주변의 환경 인자가 대칭 함수 값으로 규정되고 그 고유한 에너지를 머신 러닝으로 예측한다. 모든 원자들의 에너지 합은 DFT로 계산된 에너지와 같도록 학습되고, 이를 바탕으로 포텐셜을 만들게 된다(Figure 2). 식 (3)에서 언급된 G^2 와 G^4 는 가장 대표적인 대칭 함수이고 이 외에도 다른 대칭함수들이 활용되기도 한다[10].

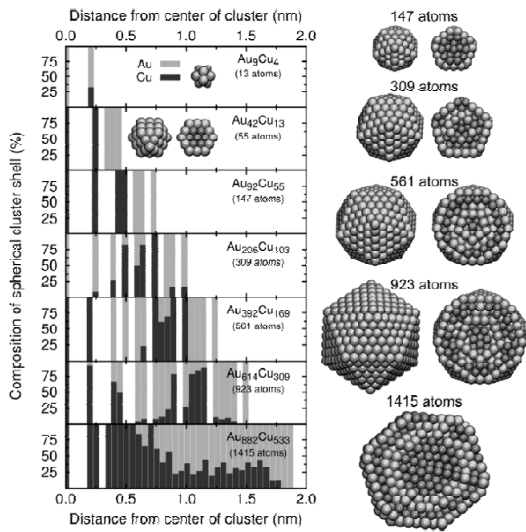
한편, 위와 같은 방법론을 바탕으로 다양한 머신 러닝 포텐셜이 발표되었는데, 대표적으로 브라

운 대학교의 A. Peterson 교수 그룹의 AMP[11], MIT의 A. M. Kolpak 교수 그룹의 PROPhet[12], N. Artrith 박사가 주도적으로 개발 중인 aenet[13], 서울대학교 한승우 교수 그룹의 SIMPLE-NN[14] 등이 있다. 또한 만들어진 머신 러닝 포텐셜은 LAMMPS를 통해서 분자 동역학 계산도 가능하다. 언급된 머신 러닝 포텐셜들은 리튬이온전지의 음극 실리콘이 리튬과 반응했을 때 형성되는 다양한 구조[15], 이종 합금 이산화탄소 환원 촉매[16], 수소 환경에서의 준안정(metastable) 상태의 백금 나노 입자 구조 예측[17] 등 다양한 분야에 적용된 사례가 있다.

3. 머신 러닝 포텐셜 적용 예시: 합금 나노 입자 구조 예측

일반적인 제일원리 DFT 계산법은 2 nm 이상의 단일 원소 나노 입자조차 많은 컴퓨팅 리소스로 장시간이 소요된다. 이종 합금 나노 입자의 경우 가능한 배열을 DFT로 스크린하여 안정한 구조를 예측한 결과들이 있지만 2 nm 이하의 크기만 고려하는 경우가 대부분이고 삼원계 합금의 경우 몇 가지 간단한 구조만 예측됐다[18,19]. 이러한 한계가 있는 상황에서 A. M. Kolpak 교수 그룹에서 머신 러닝 포텐셜을 활용해서 Au-Cu 이종 나노 합금의 6 nm 크기까지의 구조를 예측하고 더 나아가 물이 나노 입자 주변에 있을 때까지 효과를 고려한 구조를 제시했다[5]. 약 18,000개의 구조 계산 데이터를 토대로 인공 신경망을 통해 에너지를 예측한 결과, 훈련 집합에서는 8.2 meV/atom의 평균 제공근 오차를 보였고, 그 결과를 토대로 2,000개의 평가 집합에서는 12.4 meV/atom의 오차를 보였다[5]. 이는 DFT 계산 결과와 차이가 거의 없음을 시사한다. 머신 러닝 포텐셜을 이용하여 총 원자 개수가 147, 309, 561, 923, 1,415개일 때의 Au-Cu 이종 합금의 안정한 구조를 예측한 결과 Figure 3과 같이 금이 가장 바깥 표면을 형성하며 코어-셸(core-shell) 구조를 이루는 것을 확인할 수 있었다[5].

단일 또는 이종 합금 나노 입자 경우와 달리 삼

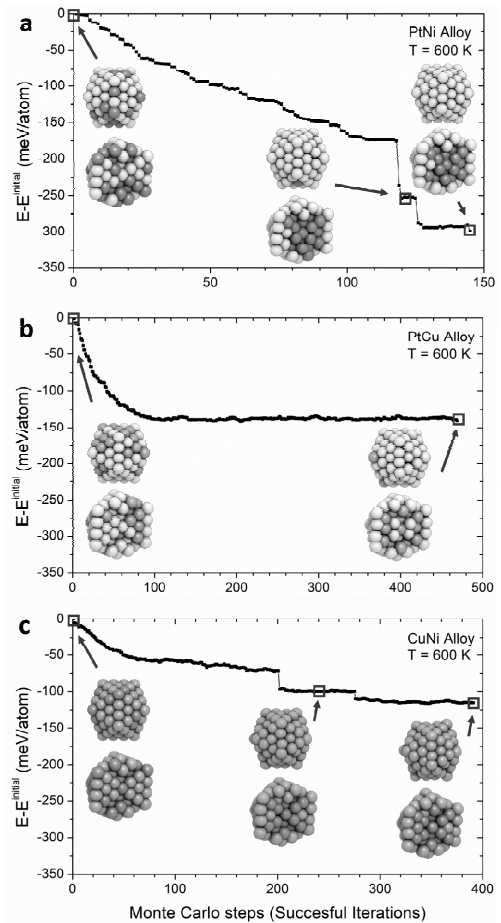


*출처: *Nano Lett.*, **14**, 2670-2676 (2014).

Figure 3. 머신 러닝 포텐셜을 통한 금-구리 이종 합금 나노 구조 예측.

성분계 이상의 경우 DFT 계산은 사실상 매우 복잡하게 생각되어 왔다. 최근 이러한 문제를 해결하고자 BPNNP 방법을 이용하여 극복한 사례가 있다[6]. 우선, 100개 이내의 2 nm 이하의 삼성분계 (Pt-Ni-Cu) 나노 입자를 학습시켜 삼성분계 합금 포텐셜 라이브러리를 개발하고, 이를 몬테카를로 방법에 입력으로 사용하여 2 nm보다 큰 나노 입자의 열역학적 구조를 찾아냈다. 이를 위해 DFT 계산 수립 전 예측되는 다양한 불안정한 구조들과 그 에너지들도 합금 포텐셜에 포함하여 머신 러닝의 정확도와 속도를 향상시켰다. 이로써, 총 2,724 (PtNi), 2,029 (PtCu), 1,554 (CuNi), 7,570 (PtCuNi) 개의 데이터를 확보하고 컷오프 반지름 6.5 Å를 사용하여 인공 신경망 훈련을 한 결과 평균 제곱근 오차는 7 meV/atom 이내의 매우 우수한 결과를 보였다[6].

이 포텐셜을 이원계 나노 합금 구조에 적용한 결과 최외각 셸(shell)은 백금 원자가 차지하고, Ni-Cu 합금에서는 Cu가 표면으로 나오는 것을 확인할 수 있었다(Figure 4). Pt과 Cu가 가장 바깥 셸로 분포하는 이유는 원자 크기와 관련이 있다. 벌크 구조의 격자 상수를 비교해보면 Pt (3.969 Å) > Cu

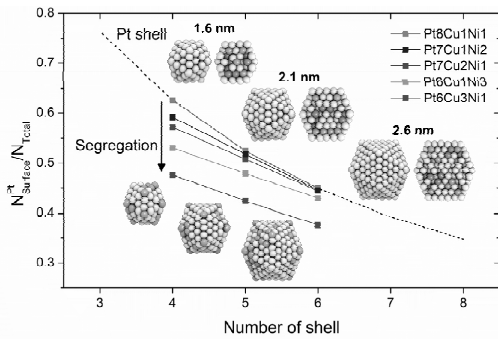


*출처: *Phys. Chem. Chem. Phys.*, **20**, 24539-24544 (2018).

Figure 4. 이종 합금 나노 구조에 대한 몬테카를로 계산 결과.

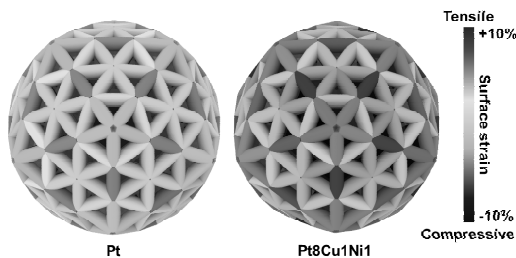
(3.627 Å) > Ni (3.518 Å) 순으로 백금이 가장 큰 것을 알 수 있다. 가장 큰 원소가 밖으로 나와야 나노 입자의 곡률로 인한 압축 변형에 유리하기 때문이다. Figure 5에 나온 것과 같이 삼원계 나노 입자에 대해서도 다양한 조성과 크기에 대해서 몬테카를로 계산을 진행했다. 바깥 표면에는 가장 큰 Pt이 나오려는 경향이 있었고, Ni 또는 Cu의 함량을 증가시키면 이들도 표면으로 석출되는 것을 확인할 수 있었다.

Figure 6은 벌크 백금의 원자간 결합 길이와 비교해서 그 변화 정도를 색으로 나타낸 것이다. 나노 입자의 표면 압축 응력은 성분 원소의 분포와 밀접한 관계가 있으며, 촉매 성능에도 영향을 주



*출처: *Phys. Chem. Chem. Phys.*, **20**, 24539-24544 (2018).

Figure 5. 기계 학습 기반 포텐셜로 예측한 삼중 합금 나노 구조.



*출처: *Phys. Chem. Chem. Phys.*, **20**, 24539-24544 (2018).

Figure 6. 백금 벌크 구조 대비 표면 결합 길이 변형을.

게 된다. 따라서 이러한 변형 정도를 파악하는 것은 촉매 소재 설계에서 고려해야 할 중요한 요인이다. 원자가 꼭짓점에 있는지 면에 있는지에 따라서 약간씩 차이는 있지만 내부에 크기가 작은 Cu 또는 Ni이 존재하면 표면의 Pt이 더 큰 압축 변형을 받는 것을 알 수 있다. 이는 순수 백금 나노 입자 대비 산소 흡착 에너지가 약 0.3 eV 약해지게 하여 산소 환원 반응 촉매 성능을 증가시키게 된다.

4. 결 론

본 기고문에서 우리는 신소재 전산 설계분야에서 제일원리와 머신 러닝을 연동하여, 정확하고 빠르게 나노 입자의 구조를 예측하는 인공 신경망 포텐셜 법에 대해 설명하고 몇 가지 예를 보였다. 대칭 함수를 이용하면 구조에 대한 정보를 컴퓨터가 인식 가능한 숫자로 표현할 수 있고 이를 토대로

인공 신경망 학습을 통해 에너지를 예측할 수 있다. 이 방법론은 기존 양자역학 기반 DFT 계산보다 100배 이상의 계산 속도를 보이며, 정확도 또한 매우 높다. 따라서 고성능/고안정을 위해 점점 복잡해지는 공업화학 소재 분야에서 새로운 물질을 설계할 때 좋은 도구가 될 수 있다. 앞으로 컴퓨터의 하드웨어/소프트웨어 성능 발전과 새로운 인공 지능 기법은 보다 효율적인 계산을 가능하게 할 것이며 전산 공업화학 분야의 중요성은 계속해서 더욱 커지게 될 것이다.

References

1. L. Tang, C. Friesen, K. Sieradzki, B. Han, G. Ceder, K. Persson, and T. He, Electrochemical stability of nanometer-scale Pt particles in acidic environments, *J. Am. Chem. Soc.*, **132**, 596-600 (2010).
2. S. H. Noh, J. Hwang, J. Kang, M. H. Seo, D. Choi, and B. Han, Tuning the catalytic activity of heterogeneous two-dimensional transition metal dichalcogenides for hydrogen evolution, *J. Mater. Chem. A*, **6**, 20005-20014 (2018).
3. S. H. Noh, C. Kwon, J. Hwang, T. Ohsaka, B. J. Kim, T. Y. Kim, Y. G. Yoon, Z. Chen, M. H. Seo, and B. Han, Self-assembled nitrogen-doped fullerenes and their catalysis for fuel cell and rechargeable metal-air battery applications, *Nanoscale*, **9**, 7373-7379 (2017).
4. J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Phys. Rev. Lett.*, **98**, 1-4 (2007).
5. N. Artrith and A. M. Kolpak, Understanding the composition and activity of electrocatalytic nanoalloys in aqueous solvents: A combination of DFT and accurate neural network potentials, *Nano Lett.*, **14**, 2670-2676 (2014).
6. J. Kang, S. H. Noh, J. Hwang, H. Chun, H.

- Kim, and B. Han, First-principles database driven computational neural network approach to the discovery of active ternary nanocatalysts for oxygen reduction reaction, *Phys. Chem. Chem. Phys.*, **20**, 24539-24544 (2018).
- W. Jeong, K. Lee, D. Yoo, D. Lee, and S. Han, Toward reliable and transferable machine learning potentials: Uniform training by overcoming sampling bias, *J. Phys. Chem. C*, **122**, 22790-22795 (2018).
 - N. Artrith, T. Morawietz, and J. Behler, High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide, *Phys. Rev. B - Condens. Matter Mater. Phys.*, **83**, 1-4 (2011).
 - M. Gastegger, J. Behler, and P. Marquetand, Machine learning molecular dynamics for the simulation of infrared spectra, *Chem. Sci.*, **8**, 6924-6935 (2017).
 - J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, *J. Chem. Phys.*, **134**, DOI: 10.1063/1.3553717.
 - A. Khorshidi, and A. A. Peterson, Amp: A modular approach to machine learning in atomistic simulations, *Comput. Phys. Commun.*, **207**, 310-324 (2016).
 - B. Kolb, L. C. Lentz, and A. M. Kolpak, Discovering charge density functionals and structure-property relationships with PROPhet: A general framework for coupling machine learning and first-principles methods, *Sci. Rep.*, **7**, 1-9 (2017).
 - N. Artrith and A. Urban, An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO₂, *Comput. Mater. Sci.*, **114**, 135-150 (2016).
 - K. Lee, D. Yoo, W. Jeong, and S. Han, SIMPLE-NN: An efficient package for training and executing neural-network interatomic potentials, *Comput. Phys. Commun.*, **242**, 95-103 (2019).
 - N. Artrith, A. Urban, and G. Ceder, Constructing first-principles phase diagrams of amorphous Li_xSi using machine-learning-assisted sampling with an evolutionary algorithm, *J. Chem. Phys.*, **148**, DOI:10.1063/1.5017661.
 - Z. W. Ulissi, M. T. Tang, J. Xiao, X. Liu, D. A. Torelli, M. Karamad, K. Cummins, C. Hahn, N. S. Lewis, T. F. Jaramillo, K. Chan, and J. K. Nørskov, Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for CO₂ reduction, *ACS Catal.*, **7**, 6600-6608 (2017).
 - G. Sun and P. Sautet, Metastable structures in cluster catalysis from first-principles: Structural ensemble in reaction conditions and metastability triggered reactivity, *J. Am. Chem. Soc.*, **140**, 2812-2820 (2018).
 - S. H. Noh, M. H. Seo, J. K. Seo, P. Fischer, and B. Han, First principles computational study on the electrochemical stability of Pt-Co nanocatalysts, *Nanoscale*, **5**, 8625-8633 (2013).
 - S. H. Noh, B. Han, and T. Ohsaka, First-principles computational study of highly stable and active ternary PtCuNi nanocatalyst for oxygen reduction reaction, *Nano Res.*, **8**, 3394-3403 (2015).

저자소개



강 준 희

2010~2012 서울시립대학교 물리학과 학사
2012~2016 대구경북과학기술원
에너지시스템공학과 박사
2016~2018 연세대학교 화학공학과
박사후연구원
2018~현재 한국에너지기술연구원
선임연구원



한 병 찬

1991~1998 서울대학교 원자핵공학과 학사
1998~2000 서울대학교 원자핵공학과 석사
2000~2007 MIT 재료공학과 박사
2007~2009 MIT 박사후 연구원
2009~2011 Stanford 대학교 박사후 연구원
2011~2014 DGIST, 조교수/부교수
2015~현재 연세대학교 화공생명공학과
교수

