

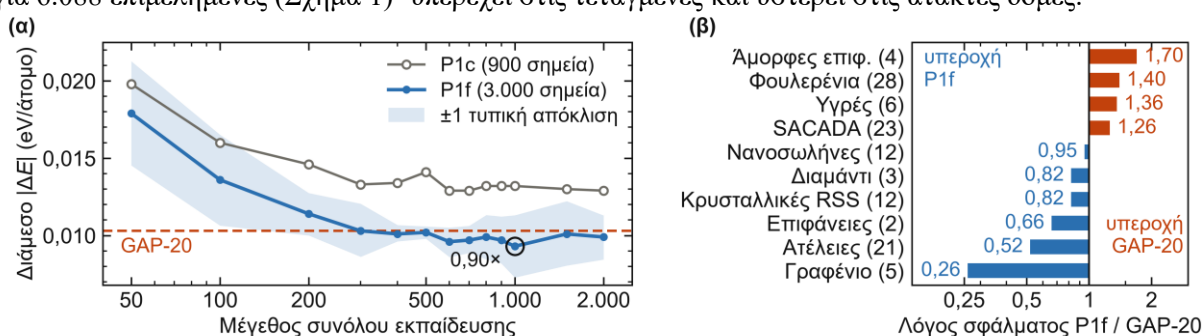
Δυναμικό GAP για τον άνθρακα από τυχαία επιλεγμένες, μη επιμελημένες δομές: σύγκριση με το GAP-20

Χ. Δ. Χαλκίδης¹, Γ. Φανουργάκης¹

¹Εργαστήριο Κβαντικής και Υπολογιστικής Χημείας, Τμήμα Χημείας, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 54124 Θεσσαλονίκη
email: chalchri@chem.auth.gr, fanourg@chem.auth.gr

Τα δυναμικά μηχανικής μάθησης, όπως τα Gaussian Approximation Potentials (GAP; Bartók et al., 2010), προσεγγίζουν την κβαντομηχανική ακρίβεια με κόστος μικρότερο κατά τάξεις μεγέθους, μεταθέτουν όμως τον περιορισμό τους από την αναλυτική μορφή στα δεδομένα, των οποίων η επαναληπτική επιμέλεια απαιτεί χρόνο και εξειδίκευση. Εξετάζεται τι ακρίβεια προκύπτει χωρίς καμία επιμέλεια και τι την περιορίζει. Σύνολα εκπαίδευσης 50 έως 2.000 δομών αντλήθηκαν με στρωματοποιημένη τυχαία δειγματοληψία από την ίδια βάση δεδομένων 17.525 διαμορφώσεων άνθρακα από την οποία προήλθε με επιμέλεια το δημοσιευμένο δυναμικό GAP-20 (Rowe et al., 2020). Εκπαιδεύτηκαν 1.040 δυναμικά με περιγραφέα SOAP (Bartók et al., 2013), σε οκτώ ρυθμίσεις με δέκα επαναλήψεις ανά μέγεθος, και αξιολογήθηκαν σε κοινό σύνολο 203 δομών.

Το σφάλμα ενέργειας μειώνεται κατά 43% έως τις 700 περίπου δομές και αυξάνεται κατά 6,6% έως τις 2.000, όχι λόγω υπερπροσαρμογής αλλά εξαιτίας σπάνιων απωστικών δομών με ακραίες δυνάμεις, που εισέρχονται συχνότερα στα μεγαλύτερα σύνολα και υποβαθμίζουν την προσαρμογή. Η δεκαπλάσια χωρητικότητα (από 300 σε 3.000 αντιπροσωπευτικά σημεία) μειώνει το σφάλμα κατά 40,5%, όχι όμως τον ρυθμό αστοχίας στις απωστικές δομές, όπου και το GAP-20 έχει σφάλμα 33 φορές μεγαλύτερο απ' ό,τι στις υπόλοιπες. Στις 116 δομές αξιολόγησης που δεν είδε ούτε το GAP-20, το καλύτερο δυναμικό φτάνει 0,0093 eV/άτομο έναντι 0,0103 του GAP-20, με 1.000 τυχαίες δομές αντί για 6.088 επιμελημένες (Σχήμα 1)· υπερέρχει στις τεταγμένες και υστερεί στις άτακτες δομές.



Σχήμα 1. (α) Διάμεσο $|\Delta E|$ στις 116 δομές που κανένα μοντέλο δεν είδε (σκίαση: ± 1 τυπική απόκλιση μεταξύ επαναλήψεων). (β) Λόγος σφάλματος P1f/GAP-20 ανά οικογένεια, στις 1.000 δομές.

Ο δεσμευτικός περιορισμός της ακρίβειας βρίσκεται επομένως στα δεδομένα, όχι στο μοντέλο· η επιμέλεια δεν είναι προαπαιτούμενο αλλά ένας από τους διαθέσιμους μοχλούς.

Αναφορές

- Bartók, A. P., Payne, M. C., Kondor, R., & Csányi, G. (2010). Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical Review Letters*, 104(13), 136403. <https://doi.org/10.1103/PhysRevLett.104.136403>
- Bartók, A. P., Kondor, R., & Csányi, G. (2013). On representing chemical environments. *Physical Review B*, 87(18), 184115. <https://doi.org/10.1103/PhysRevB.87.184115>
- Rowe, P., Deringer, V. L., Gasparotto, P., Csányi, G., & Michaelides, A. (2020). An accurate and transferable machine learning potential for carbon. *The Journal of Chemical Physics*, 153(3), 034702. <https://doi.org/10.1063/5.0005084>

Ευχαριστίες

Ευχαριστούμε το Κέντρο Ηλεκτρονικής Διακυβέρνησης ΑΠΘ για την υποδομή «Αριστοτέλης».