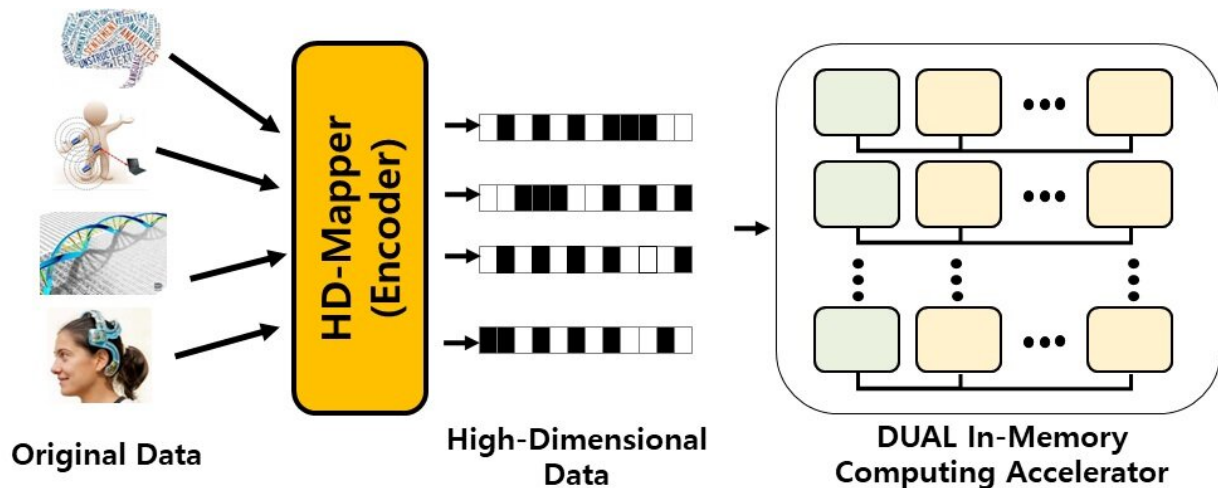


# DUAL takes AI to the next level

January 4 2021



Software and hardware overview of DUAL based on Hyperdimensional Computing. Credit: IEEE/ACM International Symposium on Microarchitecture 2020

Scientists at DGIST in Korea, and UC Irvine and UC San Diego in the US, have developed a computer architecture that processes unsupervised machine learning algorithms faster, while consuming significantly less energy than state-of-the-art graphics processing units. The key is processing data where it is stored in computer memory and in an all-digital format. The researchers presented the new architecture, called DUAL, at the 2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture.

"Today's computer applications generate a large amount of data that needs to be processed by [machine learning](#) algorithms," says Yeseong Kim of Daegu Gyeongbuk Institute of Science and Technology (DGIST), who led the effort.

Powerful "unsupervised" machine learning involves training an algorithm to recognize patterns in [large datasets](#) without providing labeled examples for comparison. One popular approach is a clustering algorithm, which groups similar data into different classes. These algorithms are used for a wide variety of data analyzes, such as identifying [fake news](#) on social media, filtering spam email and detecting criminal or fraudulent activity online.

Scientists have been looking into processing in-memory (PIM) approaches to these problems. But most PIM architectures are analog-based and require analog-to-digital and digital-to-analog converters, which take up a huge amount of the computer chip power and area. They also work better with supervised machine learning, which includes labeled datasets to train the algorithm.

To overcome these issues, Yeseong Kim of Daegu Gyeongbuk Institute of Science and Technology (DGIST) and his colleagues developed DUAL, which stands for digital-based unsupervised learning acceleration. DUAL enables computations on [digital data](#) stored inside a computer memory. It works by mapping all the data points into high-dimensional space; imagine [data points](#) stored in many locations within the human brain.

"Today's computer applications generate a large amount of data that needs to be processed by machine learning algorithms," says Kim. "But running clustering algorithms on traditional cores results in high energy consumption and slow processing, because a large amount of data needs to be moved from the computer's memory to its processing unit, where

the machine learning tasks are conducted."

The scientists found DUAL efficiently speeds up many clustering algorithms, using a wide range of large-scale datasets, and significantly improves [energy efficiency](#) compared to a state-of-the-art graphics processing unit. The researchers believe this is the first digital-based PIM architecture that can accelerate unsupervised machine learning.

"The existing approach of the state-of-the-arts in-memory computing research focuses on accelerating supervised learning algorithms through [artificial neural networks](#), which increases chip design costs and may not guarantee sufficient learning quality," says Kim. "We showed that combining hyper-dimensional and in-memory computing can significantly improve efficiency while providing sufficient accuracy."

**More information:** Mohsen Imani et al, DUAL: Acceleration of Clustering Algorithms using Digital-based Processing In-Memory, *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)* (2020). [DOI: 10.1109/micro50266.2020.00039](https://doi.org/10.1109/micro50266.2020.00039)

Provided by DGIST (Daegu Gyeongbuk Institute of Science and Technology)

Citation: DUAL takes AI to the next level (2021, January 4) retrieved 17 July 2024 from <https://techxplore.com/news/2021-01-dual-ai.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
---