Hyperon Meets Hyperdimensional Computing

Hyperon's decentralized approach to AGI represents a pivotal move, integrating diverse AI methodologies into a cohesive, shared knowledge network. Central to this integration is Hyperdimensional computing (HDC), utilizing high-dimensional spaces for nuanced information representation.

There is a convergence of Hyperon and the robust geometric approach to artificial memory afforded by Hyperdimensional computing. Hyperon's strengths are in scaling different AI approaches into a unified decentralized web of shared knowledge. This approach can afford many types of computing that resembles human cognition. To support these strengths, hyperdimensional computing (HDC) bolsters Hyperon with a unique type of learning that weaves together information in such a way that the progress of AI can rapidly evolve in a faster, more scalable, and flexible direction towards safe AGI.

The melding of hyperdimension computing into Hyperon two affords a unique synergy that naturally supports emerging intelligence. Yet, there is more to be found in thread woven between the two.

The original work of Kanerva in 1988 describes a geometric approach to artificial memory that relies on symbolic representations in high-dimensional space [1]. In its essence, there exists a collection of integer valued vectors which exist in many more planes than the traditional X,Y,Z. These vectors represent information. By moving about the many planes, or dimensions, we find that the relationship between the vectors affords learning.

In HDC, one typical level casts their data into a collection of hypervectors. Then logical analysis can be extracted by a few simple mathematical operations. When hypervectors are constructed using zeros and ones, these operations typically take the form of xor, permute, consensus sum (i.e majority), and cosine similarity [2]. There are many flavors of HDC which take on different nuances in their mechanisms of action and applications.

The original work of Kanerva was a type of hypervector use that functions similar to human memory. If one hypervector contains one piece of information, such as an image or sentence, then combining multiple hypervector has the effect of transformation the information in a sort of holographic manor. What we mean by this is that the information is now distributed amongst all parts of the vector. And due to the statistical probability of the hyperdimensional space, each hypervector has a unique trajectory through that space. In this way, information, groups of information, and groups of groups are structured in a geometric fashion. The

distances between informations, i.e. hypervectors, describes how related the data is. The relationships described by these difference are akin and function as associative memory.

There are many approaches to computing with hypervectors. The robustness of this math has been applied to computing problems such as drug discovery, seizure detection, traffic routing, food profiling, anomaly detection, electrical load forecasting, recommender systems, computing functions, factoring, language recognition, real-time classification, electrical load anomalies, etc. [3-13].

There are several features of hypervector computing that are present in every application: robustness to noise, energy efficiency, and associative learning.

Hypervectors value to Hyperon

The core value of utilizing hypervectors in Hyperon is a mechanism for representing information in such a way that the system is memory enabled at scale.

One of Hyperon's trademarks is their, "complete, scalable, and open-source" approach to a general artificial intelligence system. By using HDC, data can be represented in a particular mathematical form, a hypervector, that expresses the information contained in the data with fewer numbers. Given the data is sufficiently large, it is transformed and compacted into the hypervector form. The resulting data packet is often significantly smaller than the original data form.

The compression-like nature of transforming data to hypervector form effectively processes more information per second than the original data form. There is often very little to no loss of information if the proportions of vector size to data size and number of compute vectors are managed appropriately. The transistors of the computer hardware are then able to process more information – resulting in significant reductions of volatile and nonvolatile memory usage.

In Hyperon, we can use binary hypervectors to transmit data between AI's that make use of memory-like, contextually rich, semantic representations. The core operations that facilitate data transformation and learning also lend themselves to efficient computing. The composition of the basic operations (xor, permute, majority, and cosine distance) allow Hyperon to execute data transfer in parallel. The process is similar to auto-vectorization in which data is transformed into a compact representation collection of serial numbers which can all be computed at once. The result is accelerated speeds of data processing.

Since we are using binary values, the overhead of floating-point computer architectures can be exploited via a bit-packing technique, resulting in reduced power consumption.

Once the hypervector representations are constructed, there exist implicitly learned structure between each vector. The associations between vectors can be directly interpreted as learned knowledge of the data. When a program runs hypervectors for computation, typically a type of superposition of the bit vectors is created according the learning needs. These contextually rich, yet still compact, superposition hypervectors can act as artificial memories. As the superposition vector gets associated with more information hypervectors, the information becomes more and more distributed. Through this series of majority and xor operations, the statistical properties of the system are then used to infer the relationships between information. This is quite similar to memories in the human brain where we associate our current experiences with past experiences, building more complex concepts of the world around us as we gain experience. Thus, memories (the hypervector ones) can travel efficiently amongst decentralized AIs.

In detail, Hyperon enables many AI strategies to collaborate over a decentralized platform via the Atomspace. Some recent work indicates that by using hypervectors as standardized representation of data, the resulting data flow can be used to bridge different neural networks [14]. This is useful for two reasons. First, all data in this component of Hyperon takes the same form, a n-dimensional hypervector which follows the same compositional bitwise logic rules of all other data . This makes transfer and gluing of AI strategies in Hyperon easier to manage – resulting in efficiency.. Secondly, superposition hypervectors, the “contextual memory-like ones” can take snapshots of any memory-enabled pathway through the Hyperon network. The outcome of this approach is Hyperon can utilize a hierarchical memory structure, allowing simultaneous and concurrent computation throughout different parts of the system as needed for intelligence.

Another feature is that HDC math is functionally complete, almost any AI in Hyperon’s network can thus leverage the compact and efficient data packet we refer to as hypervectors. There exists a homomorphism between any computable program’s math and the hypervector math. Special HDC compilers help to transfer the different mathematical dialects of Hyperon between one another, ensuring productive collaboration.

The implications of Hyperon with hypervectors is a system which can be infused with safeguards. There is a more philosophical perspective on safeguarding AI by embodying the principles of neuroscience and biology. While Hyperon is not a strictly bio-inspired AI approach, the addition of the bio-inspired HDC can be an advantage. The human brain is quite robust to damage, adapting to change, and managing its internal resources across a lifetime of learning. HDC, in the way in which it can be modularly enabled and controlled by the decentralized Hyperon system, affords some essence of these same principles and it could be a key to inner and outer alignment of safe human:AGI values.

HDC on Current Hardware

Research has demonstrated the widespread abilities of computing on current hardware. Many components of Hyperon are able to enable an HDC-memory component on current CPUs and GPUs. techniques such as: associative memory, few shot learning, regression, forecasting, time-series analysis, machine learning, unsupervised learning, memory augmented neural networks, etc. [4,15-21]. Most relevant to Hyperon is the use of HDC for IoT applications [22]. The outcome of applying HDC to AI strategies is reduced model size, fewer training iterations needed, reduced size in data needed to learn, reduced or eliminated pre-processing of data [23-36].

There are some notable works in which HDC increases speech recognition with 11.9x higher energy efficiency, 5.3 faster inference and 4.6 faster training, DNA sequencing at 69% less energy consumption and 77I % faster, DNA pattern matching with 26x speedup and 40x higher energy efficiency, human activity recognition was 486x faster with 7x lower power [11,25,28]. Thus, infusing AI with HDC already holds powerful benefits in making AI more efficient, while providing the same or better accuracy.

Specialized Hypervector Hardware

The value of increased speed, reduced memory load, and overall power reduction is orders of magnitude larger in custom hardware. There are recent works that demonstrate the acceleration of this technology in academic prototypes [29,36-41]. Their results on average suggest performance gains of 642x speedup and 583x increased energy efficiency during AI training over CPUs — 19x speedup and 107x energy efficiency increase over GPUs. Specialized hardware reduces the computational complexity for HDC [22].

The most critical gain opportunity that specialized hardware for HDC affords is speed. For example, the most commonly used operation, xor, optimized on Intel's AVX architectures computes the output of two 8,192 dimensional hypervectors in 9 microseconds. On an FPGA prototype, we've reduced this time to 2.4 microseconds. For a speedup of 3.75x on a single operation that is used to encode a 4k image with 8.3 million pixels, xor is used approximately 25 million times, this is a huge improvement. Many other of the core operations in HDC follow similar trends in performance gains by directly optimizing them in circuits. Based on early performance of a single image transformation, prototyping shows a total 28x improvement in speed, 2.5x reduction in Watts, and 39x reduction in bytes used when compared to CPU and 1.7 times faster than GPU implementation.

Current hardware is reliant on floating-point architectures that have overhead related to the complexity of the core operations needed. However, in a specialized hypervector chip, the few core operations can be By exploiting the simple core operations of HDC, specialized hardware has been developed to best utilize the natural efficiencies of HDC. By stripping out the overhead of complicated architectures and functions in general purpose CPUs and floating-point SIMD GPUs, specialized hardware for HDC allows a chip that is optimized not only for speed, but also power efficiency.

The hierarchical nature of hypervector computing can be optimized to handle the large vector movement at precise steps in the Hyperon network by utilizing an MIMD structure. There is also near memory fabric that supports more efficient shuffling of the special high-dimensional bit vectors.

There is a pseudo-quantum perspective of HDC in which the values of the hypervectors behave in hyperdimensional planes of the superposition vector as qubits. By back of the envelope calculations, the point at which transistors can not get any smaller due to physical interactions of atomic particles is in 9 years. At that point, whatever advancements in the realm of transistor computing will be in the realm of quantum computing. The pseudo-quantum nature of HDC boards well in the event of this hypothetical, yet likely, transition. Hyperon will be able to transition with grace to the next major leap in computing.

References

1. Denning, P. J. (1989). *Sparse distributed memory*. Research Institute for Advanced Computer Science NASA Ames Research Center.
2. Kanerva, P. (2022). Hyperdimensional Computing: An algebra for computing with vectors. *Advances in Semiconductor Technologies*, 25–42. https://doi.org/10.1002/9781119869610.ch2
3. Bandaragoda, T., De Silva, D., Kleyko, D., Osipov, E., Wiklund, U., & Alahakoon, D. (2019). Trajectory clustering of road traffic in urban environments using incremental machine learning in combination with hyperdimensional computing. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. https://doi.org/10.1109/itsc.2019.8917320
4. Burrello, A., Schindler, K., Benini, L., & Rahimi, A. (2018). One-shot learning for IEEG seizure detection using end-to-end binary operations: Local binary patterns with hyperdimensional computing. *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. https://doi.org/10.1109/biocas.2018.8584751
5. Frady, E. P., Kleyko, D., Kymn, C. J., Olshausen, B. A., & Sommer, F. T. (2022). Computing on functions using randomized vector representations (in brief).

Neuro-Inspired Computational Elements Conference. https://doi.org/10.1145/3517343.3522597
6. Guo, Y., Imani, M., Kang, J., Salamat, S., Morris, J., Aksanli, B., Kim, Y., & Rosing, T. (2021). HyperRec. *Proceedings of the 26th Asia and South Pacific Design Automation Conference*. https://doi.org/10.1145/3394885.3431553
7. Hernández-Cano, A., Zhuo, C., Yin, X., & Imani, M. (2021). Real-time and robust hyperdimensional classification. *Proceedings of the 2021 on Great Lakes Symposium on VLSI*. https://doi.org/10.1145/3453688.3461749
8. Li, H., Wu, T. F., Rahimi, A., Li, K.-S., Rusch, M., Lin, C.-H., Hsu, J.-L., Sabry, M. M., Eryilmaz, S. B., Sohn, J., Chiu, W.-C., Chen, M.-C., Wu, T.-T., Shieh, J.-M., Yeh, W.-K., Rabaey, J. M., Mitra, S., & Wong, H.-S. P. (2016). Hyperdimensional computing with 3D VRRAM in-memory kernels: Device-architecture co-design for energy-efficient, error-resilient language recognition. *2016 IEEE International Electron Devices Meeting (IEDM)*. https://doi.org/10.1109/iedm.2016.7838428
9. Ma, D., Thapa, R., & Jiao, X. (2022). MoleHD: Efficient Drug Discovery Using Brain inspired hyperdimensional computing. *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. https://doi.org/10.1109/bibm55620.2022.9995708
10. Shahroodi, T., Zahedi, M., Firtina, C., Alser, M., Wong, S., Mutlu, O., & Hamdioui, S. (2022). Demeter: A fast and energy-efficient food profiler using hyperdimensional computing in memory. *IEEE Access*, *10*, 82493–82510. https://doi.org/10.1109/access.2022.3195878
11. Kim, Y., Imani, M., & Rosing, T. S. (2018). Efficient human activity recognition using hyperdimensional computing. *Proceedings of the 8th International Conference on the Internet of Things*. https://doi.org/10.1145/3277593.3277617
12. Singh, K. (2022). *Review of: "Multi-Centroid Hyperdimensional Computing Approach for Epileptic Seizure Detection."* https://doi.org/10.32388/oc5n6s
13. Wang, X., Yao, Z., & Papaefthymiou, M. (2023). A real-time electrical load forecasting and unsupervised anomaly detection framework. *Applied Energy*, *330*, 120279. https://doi.org/10.1016/j.apenergy.2022.120279
14. Sutor, P., Yuan, D., Summers-Stay, D., Fermuller, C., & Aloimonos, Y. (2022). Gluing neural networks symbolically through hyperdimensional computing. *2022 International Joint Conference on Neural Networks (IJCNN)*. https://doi.org/10.1109/ijcnn55064.2022.9892622
15. Khaleghi, B., Xu, H., Morris, J., & Rosing, T. S. (2021). Tiny-HD: Ultra-efficient hyperdimensional computing engine for IOT Applications. *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. https://doi.org/10.23919/date51398.2021.9473920
16. Hernandez-Cano, A., Zhuo, C., Yin, X., & Imani, M. (2021). RegHD: Robust and efficient regression in hyper-dimensional learning system. *2021 58th ACM/IEEE*

Design Automation Conference (DAC). https://doi.org/10.1109/dac18074.2021.9586284
17. Morris, J., Ergun, K., Khaleghi, B., Imani, M., Aksanli, B., & Simunic, T. (2022). HyDREA: Utilizing hyperdimensional computing for a more robust and efficient machine learning system. *ACM Transactions on Embedded Computing Systems*, *21*(6), 1–25. https://doi.org/10.1145/3524067
18. Osipov, E., Kahawala, S., Haputhanthri, D., Kempitiya, T., Silva, D. D., Alahakoon, D., & Kleyko, D. (2022). Hyperseed: Unsupervised learning with vector symbolic architectures. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15. https://doi.org/10.1109/tnnls.2022.3211274
19. Karunaratne, G., Schmuck, M., Le Gallo, M., Cherubini, G., Benini, L., Sebastian, A., & Rahimi, A. (2021). Robust high-dimensional memory-augmented neural networks. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-22364-0
20. Schlegel, K., Mirus, F., Neubert, P., & Protzel, P. (2021). Multivariate time series analysis for driving style classification using neural networks and hyperdimensional computing. *2021 IEEE Intelligent Vehicles Symposium (IV)*. https://doi.org/10.1109/iv48863.2021.9576028
21. Kleyko, D., Rachkovskij, D., Osipov, E., & Rahimi, A. (2023). A survey on hyperdimensional computing aka Vector symbolic architectures, part II: Applications, cognitive models, and challenges. *ACM Computing Surveys*, *55*(9), 1–52. https://doi.org/10.1145/3558000
22. Imani, M., Salamat, S., Gupta, S., Huang, J., & Rosing, T. (2019). Fach. *Proceedings of the 24th Asia and South Pacific Design Automation Conference*. https://doi.org/10.1145/3287624.3287667
23. Alonso, P., Shridhar, K., Kleyko, D., Osipov, E., & Liwicki, M. (2021). HyperEmbed: Tradeoffs between resources and performance in NLP tasks with hyperdimensional computing enabled embedding of N-gram statistics. *2021 International Joint Conference on Neural Networks (IJCNN)*. https://doi.org/10.1109/ijcnn52387.2021.9534359
24. Asgarinejad, F., Thomas, A., & Rosing, T. (2020). Detection of epileptic seizures from surface EEG using hyperdimensional computing. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. https://doi.org/10.1109/embc44109.2020.9175328
25. Imani, M., Kong, D., Rahimi, A., & Rosing, T. (2017). VoiceHD: Hyperdimensional Computing for efficient speech recognition. *2017 IEEE International Conference on Rebooting Computing (ICRC)*. https://doi.org/10.1109/icrc.2017.8123650
26. Imani, M., Morris, J., Bosch, S., Shu, H., Micheli, G. D., & Rosing, T. (2019). AdaptHD: Adaptive efficient training for brain-inspired hyperdimensional computing. *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. https://doi.org/10.1109/biocas.2019.8918974

27. Imani, M., Nassar, T., Rahimi, A., & Rosing, T. (2018). HDNA: Energy-efficient DNA sequencing using hyperdimensional computing. *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. https://doi.org/10.1109/bhi.2018.8333421
28. Kim, Y., Imani, M., Moshiri, N., & Rosing, T. (2020). GenieHD: Efficient DNA pattern matching accelerator using hyperdimensional computing. *2020 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. https://doi.org/10.23919/date48585.2020.9116397
29. Morris, J., Ergun, K., Khaleghi, B., Imani, M., Aksanli, B., & Rosing, T. (2021). Hydrea: Towards more robust and efficient machine learning systems with hyperdimensional computing. *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. https://doi.org/10.23919/date51398.2021.9474218
30. Morris, J., Imani, M., Bosch, S., Thomas, A., Shu, H., & Rosing, T. (2019). CompHD: Efficient hyperdimensional computing using model compression. *2019 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*. https://doi.org/10.1109/islped.2019.8824908
31. Nunes, I., Heddes, M., Givargis, T., Nicolau, A., & Veidenbaum, A. (2022). GraphHD: Efficient Graph Classification using hyperdimensional computing. *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. https://doi.org/10.23919/date54114.2022.9774533
32. Rahimi, A., Kanerva, P., & Rabaey, J. M. (2016). A robust and energy-efficient classifier using brain-inspired hyperdimensional computing. *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*. https://doi.org/10.1145/2934583.2934624
33. Rahimi, A., Kanerva, P., Benini, L., & Rabaey, J. M. (2019). Efficient Biosignal processing using hyperdimensional computing: Network templates for combined learning and classification of ExG Signals. *Proceedings of the IEEE*, *107*(1), 123–143. https://doi.org/10.1109/jproc.2018.2871163
34. Rahimi, A., Tchouprina, A., Kanerva, P., Millán, J. del, & Rabaey, J. M. (2017). Hyperdimensional computing for blind and one-shot classification of EEG error-related potentials. *Mobile Networks and Applications*, *25*(5), 1958–1969. https://doi.org/10.1007/s11036-017-0942-6
35. Thapa, R., Lamichhane, B., Ma, D., & Jiao, X. (2021). SpamHD: Memory-efficient text spam detection using brain-inspired hyperdimensional computing. *2021 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. https://doi.org/10.1109/isvlsi51109.2021.00026
36. Zou, Z., Alimohamadi, H., Zakeri, A., Imani, F., Kim, Y., Najafi, M. H., & Imani, M. (2022). Memory-inspired spiking hyperdimensional network for robust online learning. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-11073-3

37. Gupta, S., Morris, J., Imani, M., Ramkumar, R., Yu, J., Tiwari, A., Aksanli, B., & Rosing, T. Š. (2020). Thrifty. *Proceedings of the 39th International Conference on Computer-Aided Design*. https://doi.org/10.1145/3400302.3415723
38. Imani, M., Zou, Z., Bosch, S., Rao, S. A., Salamat, S., Kumar, V., Kim, Y., & Rosing, T. (2021). Revisiting hyperdimensional learning for FPGA and low-power architectures. *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. https://doi.org/10.1109/hpca51647.2021.00028
39. Kang, J., Khaleghi, B., Kim, Y., & Rosing, T. (2022). XCelHD: An efficient GPU-powered hyperdimensional computing with parallelized training. *2022 27th Asia and South Pacific Design Automation Conference (ASP-DAC)*. https://doi.org/10.1109/asp-dac52403.2022.9712549
40. Khaleghi, B., Salamat, S., Thomas, A., Asgarinejad, F., Kim, Y., & Rosing, T. (2020). SHEARer: highly-efficient hyperdimensional computing by software-hardware enabled multifold approximation. *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design*. https://doi.org/10.1145/3370748.3406587
41. Salamat, S., Imani, M., Khaleghi, B., & Rosing, T. (2019). F5-HD. *Proceedings of the 2019 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*. https://doi.org/10.1145/3289602.3293913
42.
43.