Molecular Search with the Gemini® Associative Processing Unit

Introduction

Drug discovery is expensive—both in terms of cost and time. Searching large chemical fingerprint databases requires substantial computing power. The ability to search quickly and effectively has a direct impact on time-to-market goals.

Introducing a new molecule search engine powered by GSI's Gemini® Associative Processing Unit (APU). Gemini is a patented processing technology featuring massive parallel data processing, compute and search; in-place, directly in the memory array. This architecture gives the APU a performance edge in the acceleration of similarity search applications.

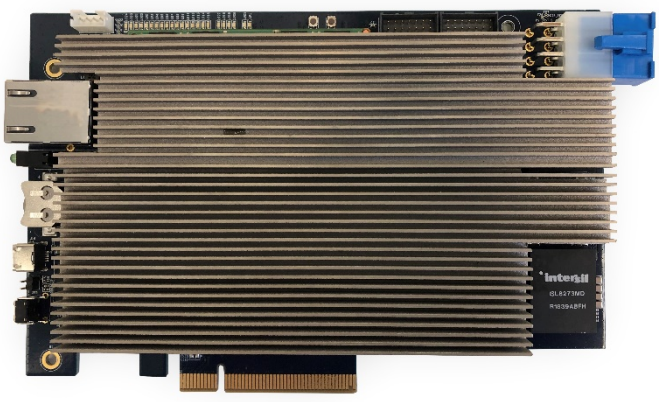
Integrated with the BIOVIA™ Pipeline Pilot scientific application, Gemini® delivers exceptional performance for circular fingerprint similarity search, including threshold searches at under 0.8.

Gemini® is a scalable solution. Large datasets and multiple queries can be processed quickly with no loss of performance and with very low power consumption. Integration with Pipeline Pilot or other applications using the Tanimoto distance metric, is straightforward.

Joining the Fight Against Covid-19

Researchers are now using Gemini® APU to significantly accelerate similarity structure searches for active chemical compounds that can target the coronavirus.

GSI APU Board with Gemini® APU



Board Specifications	
Thermal Design Power (TDP)	60W*
PCI IDs	Serial controller: GSI Technology Gemini® (Leda-G®)
APU clocks	400 MHz
PCI express interface	Gen 3 x 8
Board dimensions	254 x 111.15 mm

*40W power consumption at 400 MHz tested for Molecular Search

Spec.	One Board	Multiple Boards
Maximum memory clock	2666 MHz	2666 MHz
Memory size	16 GB	N x 16 GB
Memory bus width	64 bits	N x 64 bits
Peak memory bandwidth	21.3 GB/s	N x 21.3 GB/s

Performance and Power Consumption

The table to the left shows Gemini® APU query processing time for KNN similarity search, with 512-bit and 1024-bit vectors, 3 database sizes, and number of queries ranging from 1 to 100.

The table to the right shows search time performance on a 38M database, using the Tanimoto threshold. The threshold range is 1 – 0.01, where 1 means an exact match.

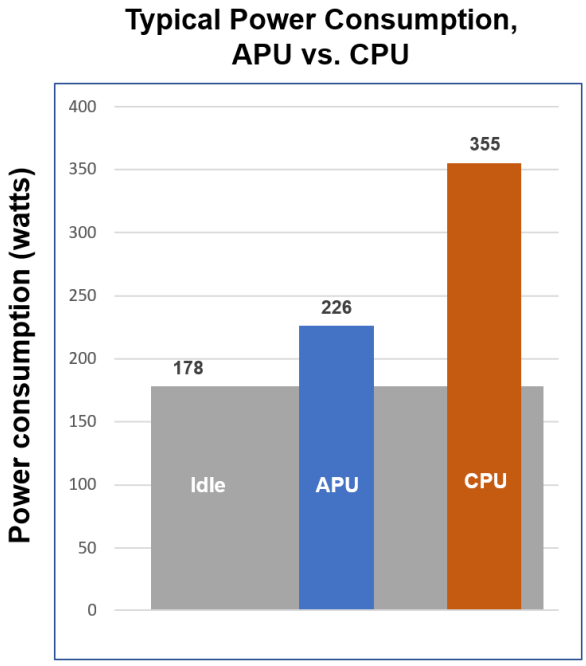
nbit	Number of Queries	Processing Time (sec)		
		10M	38M	680M (4 APUs)
512	1	0.124	1.1	4.5
	10	0.127	1.11	4.52
	50	0.129	1.18	4.55
	100	0.13	1.41	4.6
1024	1	0.25	2.11	4.52
	10	0.258	2.12	4.54
	50	0.261	2.2	4.57
	100	0.264	2.84	4.62

512-bit vectors	
Max time (sec)	1.0917
Min time (sec)	1.0025

1024-bit vectors	
Max time (sec)	2.0639
Min time (sec)	1.9853

Gemini® APU Benefits

- Support for similarity structure search on circular fingerprints at thresholds of 0.8 and under.
- Easy integration with BIOVIA—proven and fully tested.
- Gemini® can be integrated into other molecule search solutions that use the Tanimoto coefficient.
- Support for single or multiple queries on Enamine REAL database (680M compounds), all on a single APU.
- Calculation of more descriptive, larger length, folded fingerprints (bit sizes of 256, 512, 1024, and up to 8192).
- Multiple databases can be loaded into the APU's system memory. Single dataset can be chosen for a specific query.
- Batches of compounds (100s or 1000s) can be submitted simultaneously.



The APU reduces search power consumption by a factor of 3.5.

For more information on the Gemini® APU, please visit www.gsitechnology.com/APU.