A Perspective on Digital Signal Processor Based Leadership Performance Accelerator for AI and HPC

Yang GUO, Yaohua WANG, Sheng MA

Frontiers of Computer Science, DOI: 10.1007/s11704-024-40149-8

Problems & Ideas

- Existing Problems:
  - High-performance computing (HPC) and artificial intelligence (AI) require immense computational power. Traditional processors struggle with issues like high power consumption, slow data communication, and limited storage. As Moore's Law slows down, new architectures and computing methods are needed.
- Original Ideas:
  - This paper presents a DSP-based solution, highlighting three generations of FT-Matrix DSPs (FT-Matrix1, FT-Matrix2, MT-3000). It shows how to boost HPC and AI performance through techniques like vector-SIMD pipelines, better memory access, and improved data transfer.

Three generations of FT-Matrix DSPs

Future & Conclusions

- Future Directions:
  - Chiplet Technology
  - Heterogeneous Fusion Architecture:
  - Software and Hardware Codesign
  - New Computing Paradigms
  - Open Source Ecosystem Development

- Conclusion:
  - Both HPC and AI applications are motivating the development of ultra-high-performance processors with low power consumption. The combination of hardware and software ecosystems is essential for maximizing processor performance. This paper highlights the advances in DSP technology for HPC and AI domains, along with promising research directions.