1. Problem Statement: Redesigning the Datacenter for Intelligent Personal Assistants

Intelligent Personal Assistants (IPAs) are standard in today’s mobile devices. The rapid rise in IPA equipped devices means more compute intensive queries will be hitting current datacenters which are ill-suited to handle this type of workload.

2. Sirius: An Open End-to-End Voice and Vision Personal Assistant

Sirius: built from the latest open source tools; Sirius resembles current production intelligent personal assistants in its algorithmic components.

3. Accelerating Sirius-suite

Sirius-suite: extracted from Sirius, this is a suite of the 7 most computationally demanding kernels in Sirius representing 92% of the total execution time.

4. Implications for Future Warehouse Scale Computers

Latency Reduction:
FPGA: 16x
GPU: 10x

Total Cost of Ownership (TCO) Improvement
GPU: 2.6x
FPGA: 1.4x