

Cosine similarity metric calculation on low power heterogeneous computing platform

Michał Karwatowski^{1,2}, Sebastian Koryciak^{1,2},
Ernest Jamro^{1,2}, Agnieszka Dąbrowska-Boruch^{1,2},
Kazimierz Wiatr¹

¹AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków,

²ACK Cyfronet AGH, ul. Nawojki 11, 30-950 Kraków

KUKDM 11-13.03.2015 Zakopane

Agenda

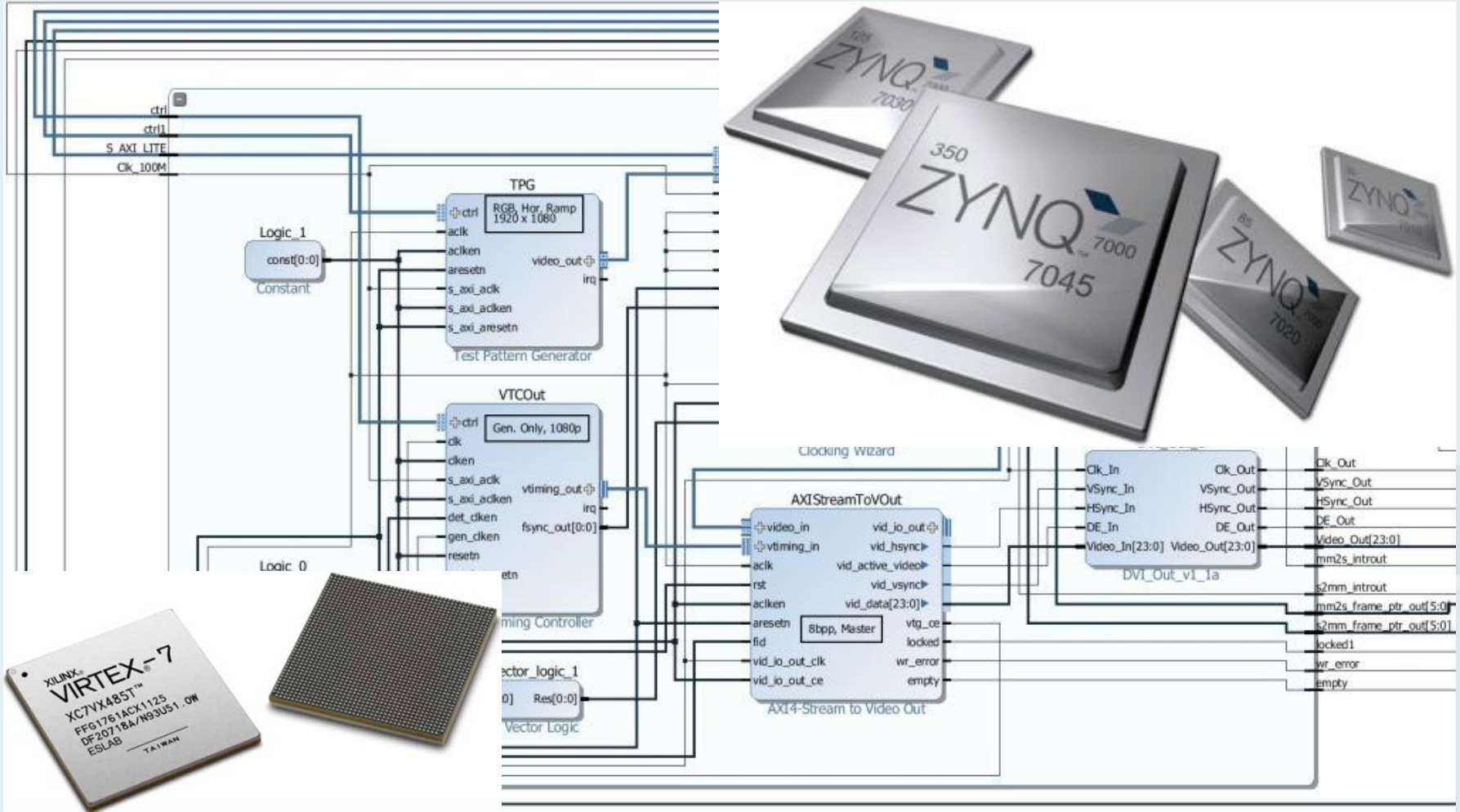


2

- FPGA based hardware accelerated computing
- Text similarity analysis
- Search algorithm implementation
- Results
- Future work



FPGA based hardware accelerated computing



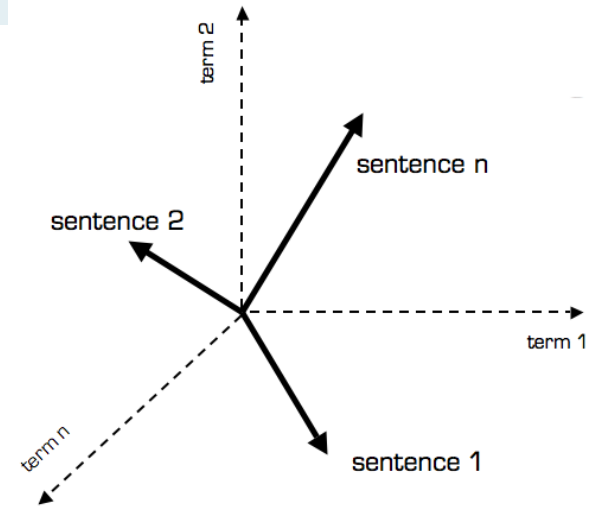
Vector Space Model

Term Frequency – Inverse Document Frequency

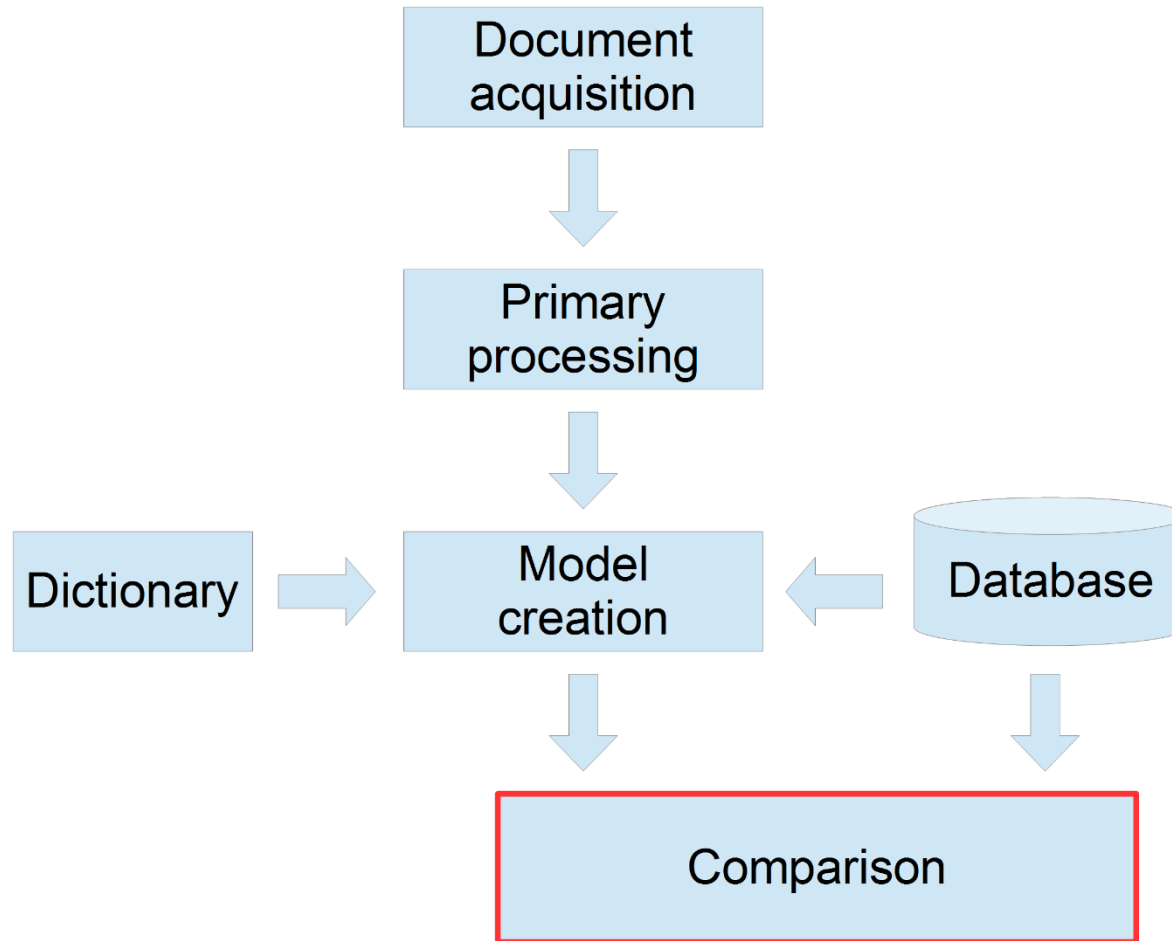
$$(tf - idf)_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \times \log \frac{|D|}{|\{d: t_i \in d\}|}$$

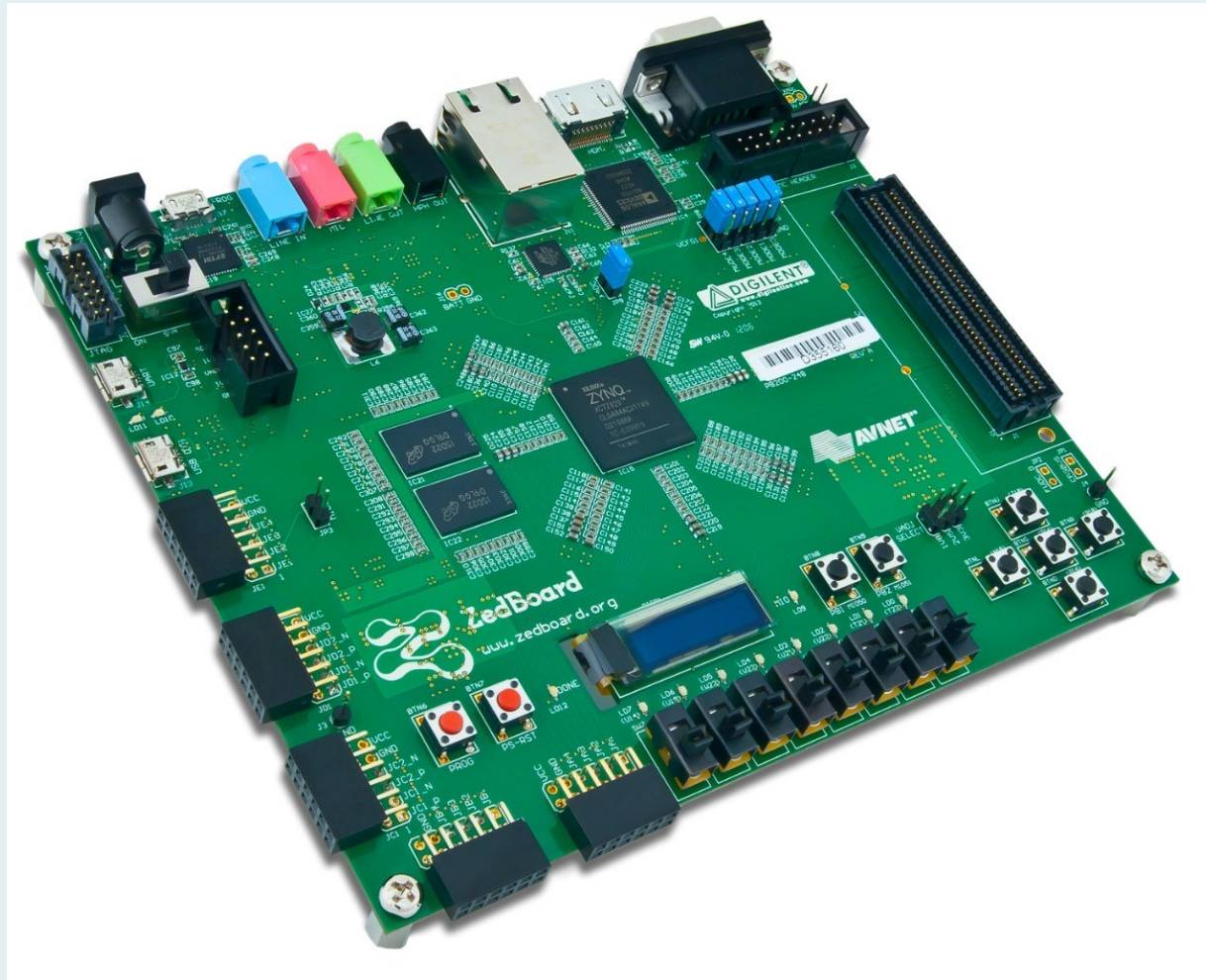
Cosine similarity

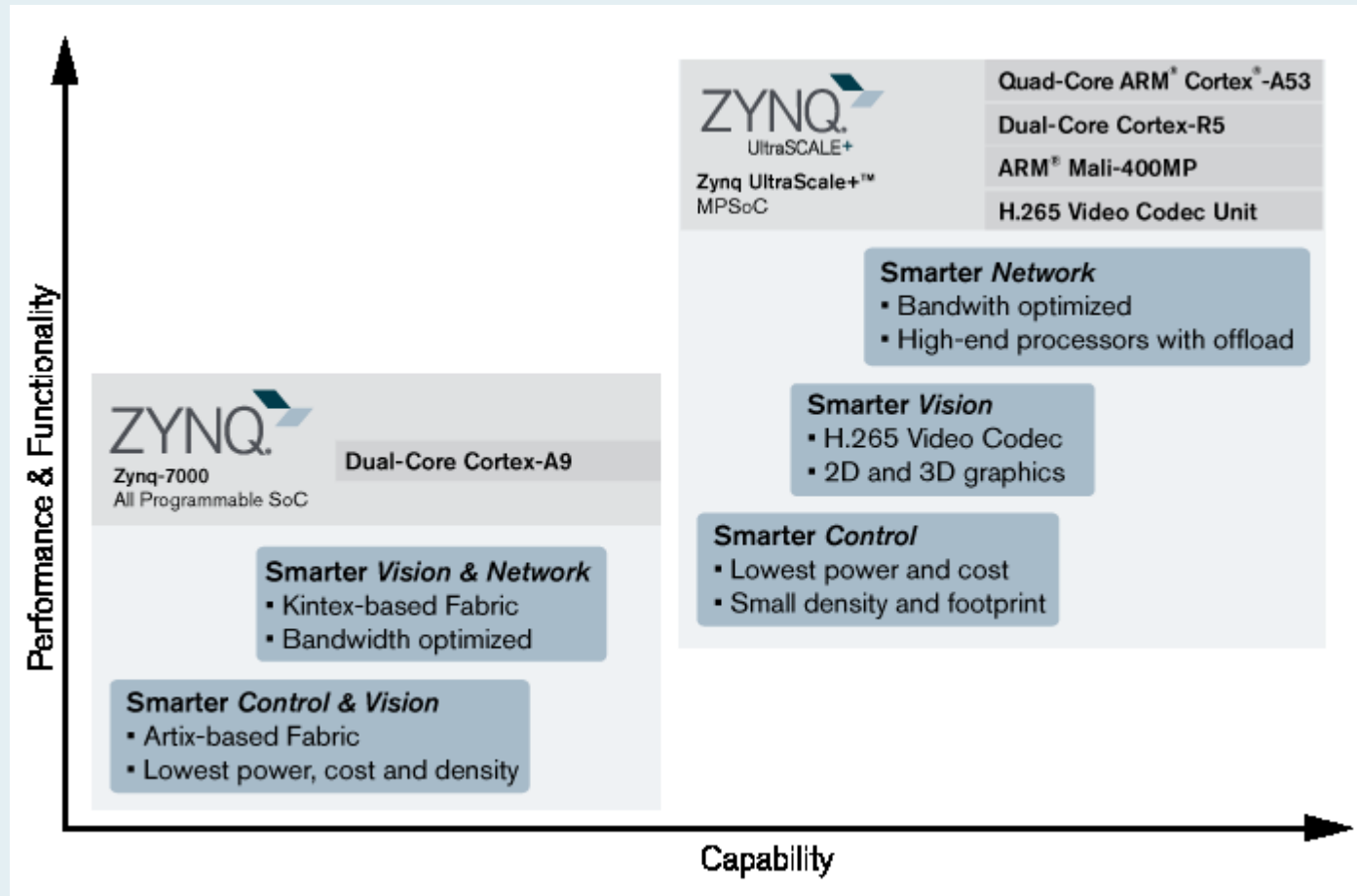
$$\cos \theta = \frac{A \cdot B}{\|A\| \|B\|}$$



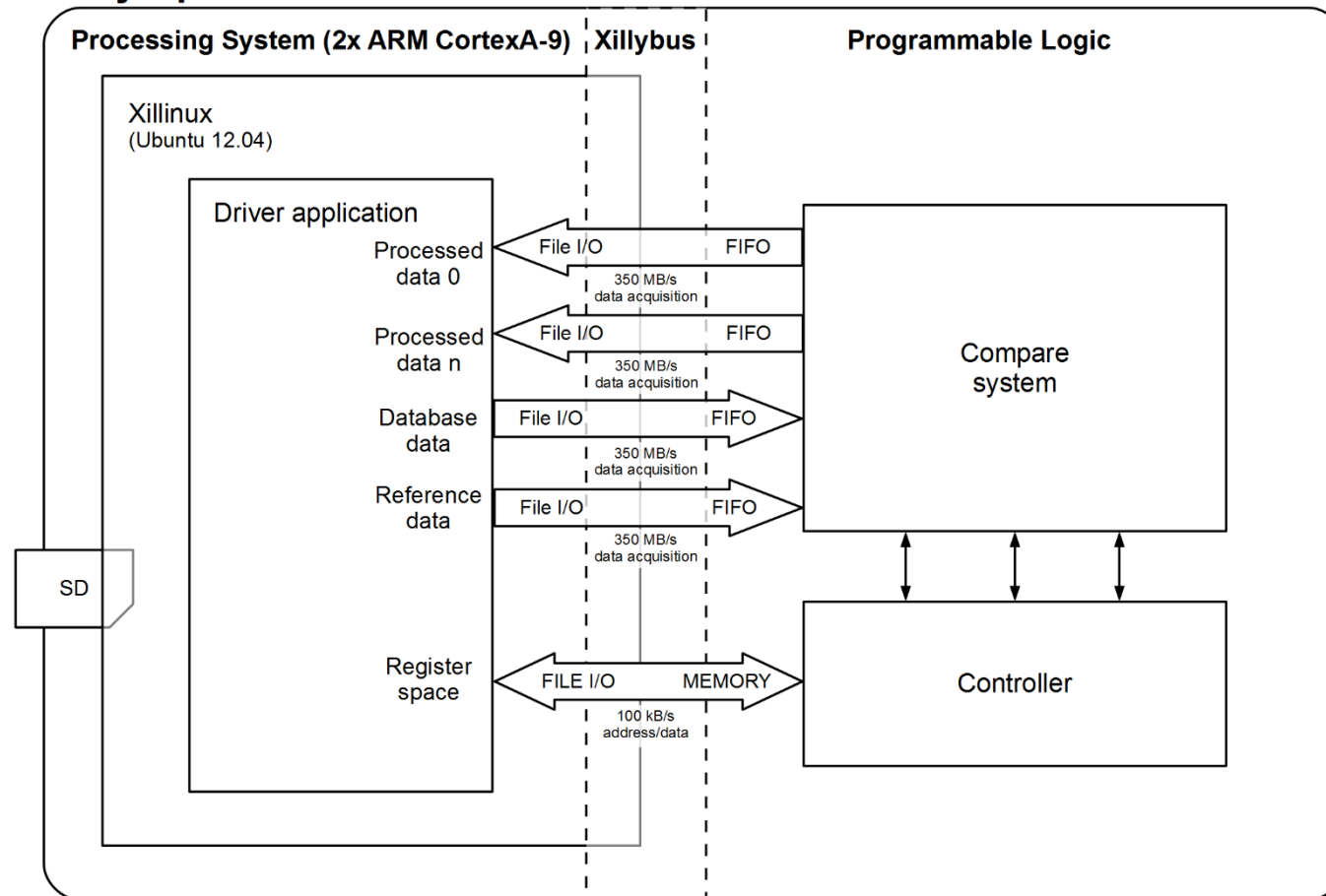
Text comparison



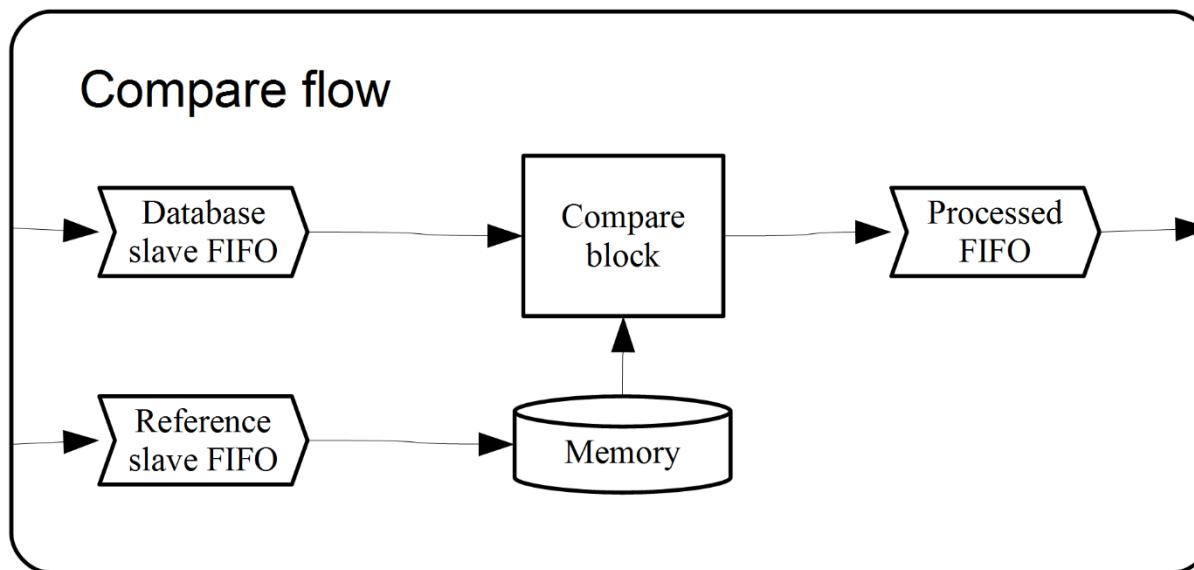




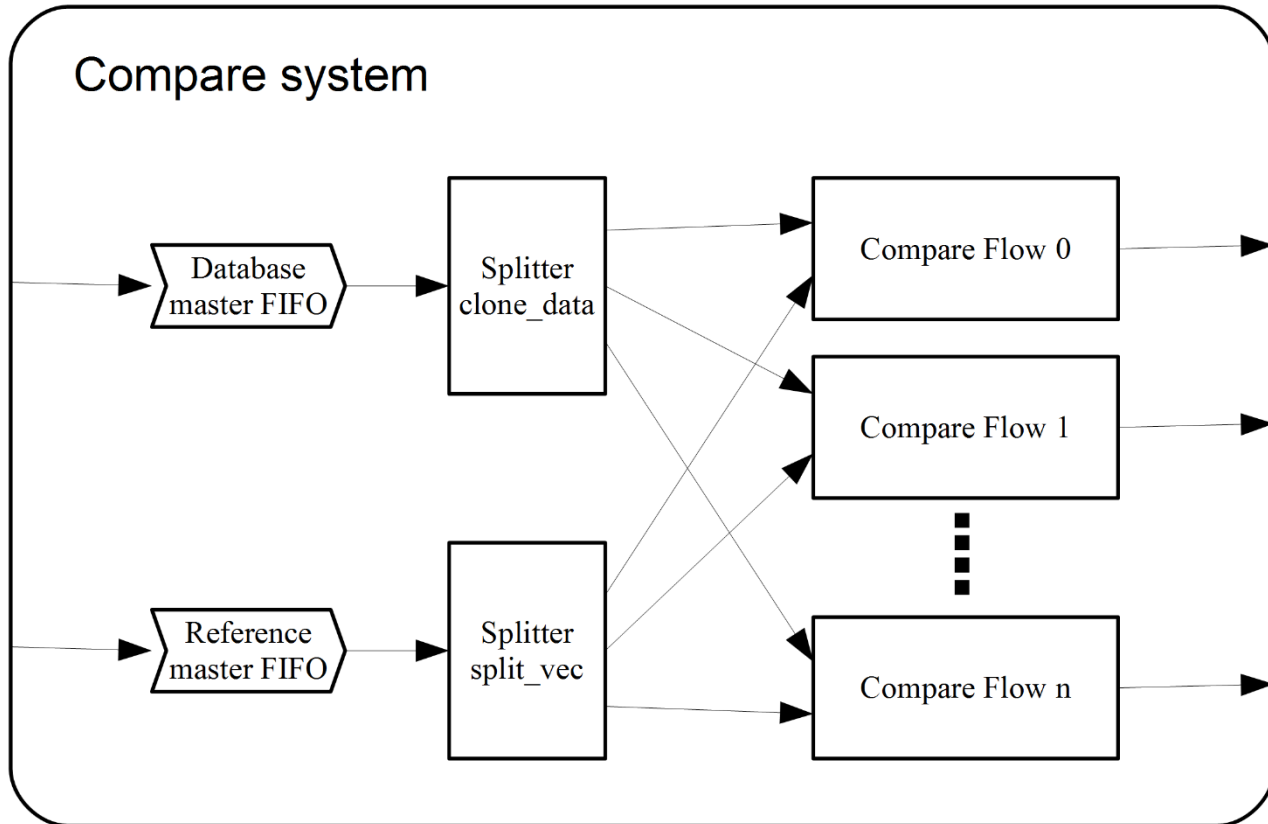
Zynq



Compare flow

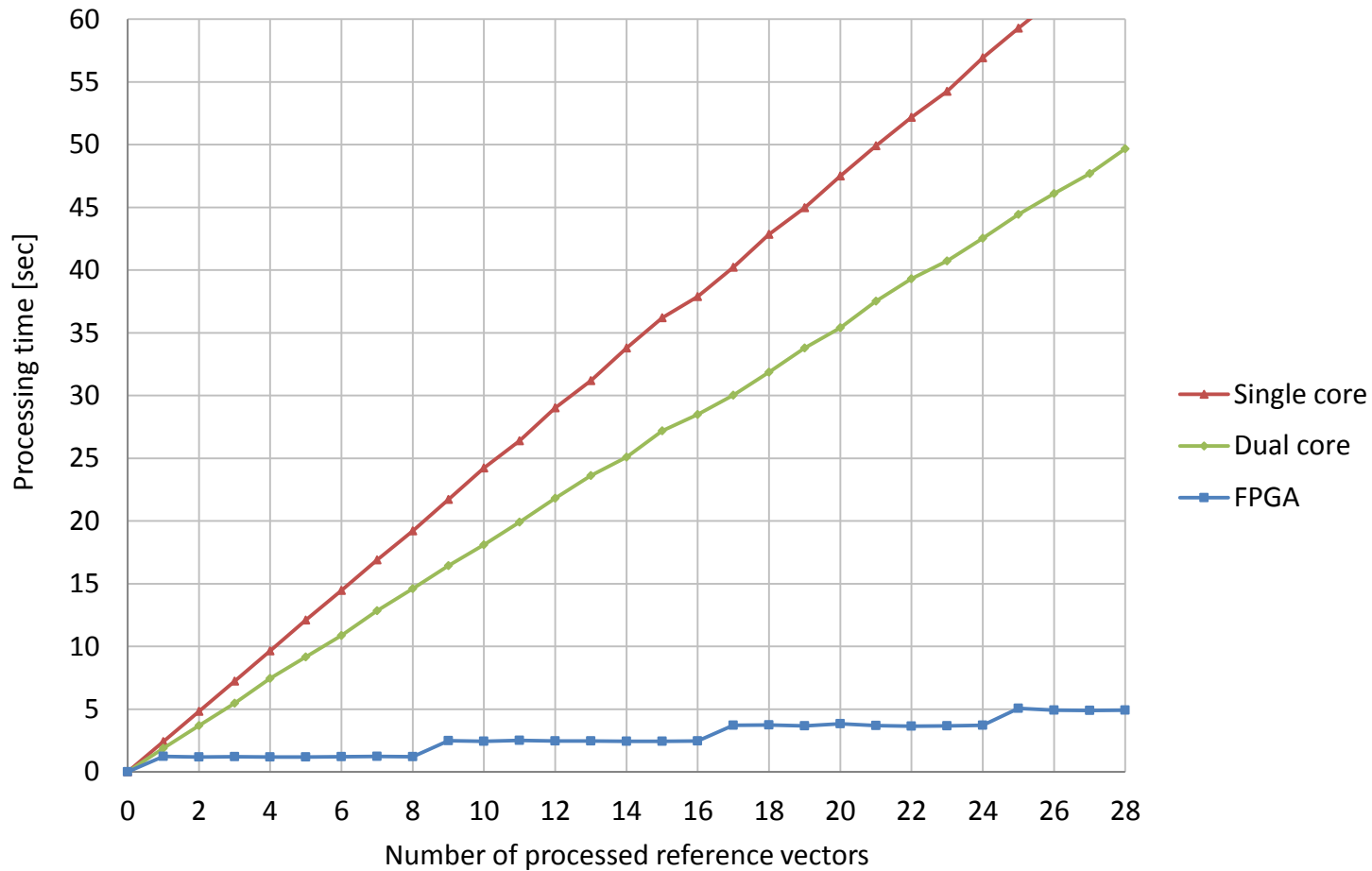


Compare system



- 100,000 random documents processed to vector form
- Zynq software solution:
 - One and two cores
 - ARM Cortex-A9
 - 667 MHz
- Zynq PS + PL solution
 - 8 paralel channels
 - 100 MHz

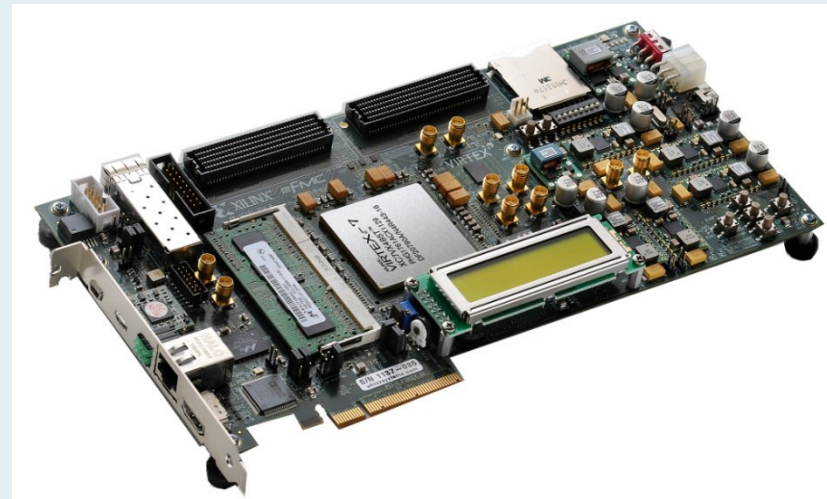
Runtime comparison



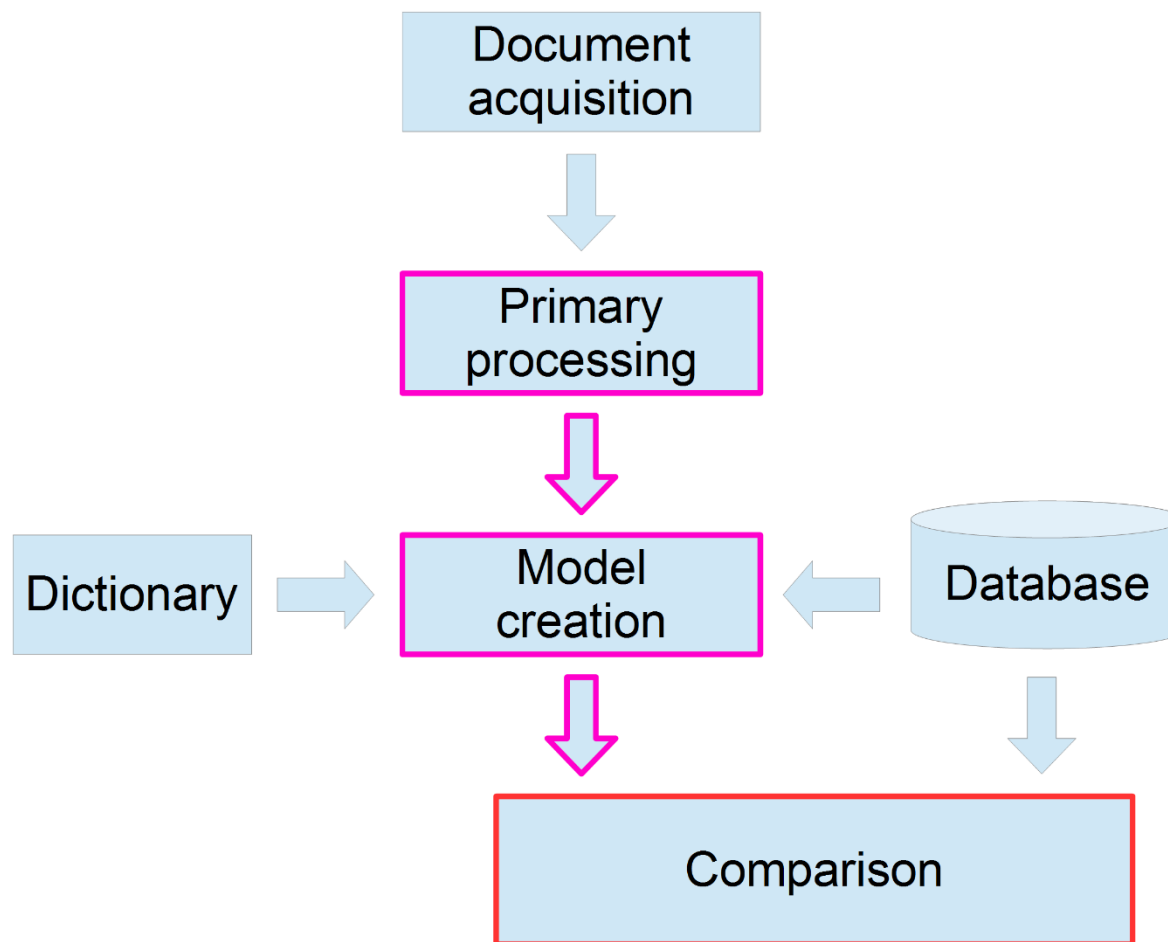
- Compression
- High performance hardware
- Higher level language



OpenCL



Text comparison



Cluster



Questions

