A Hybrid Memory Sub-system for Video Coding Applications

Su-Shin Ang [†1], George Constantinides [†], Wayne Luk [*], Peter Cheung [†]

[†]Department of Electrical and Electronics Engineering, Imperial College
South Kensington campus, London SW7 2AZ, United Kingdom
[1]su-shin.ang04@imperial.ac.uk

[*]Department of Computing, Imperial College
180 Queen's Gate, South Kensington Campus, London SW7 2AZ, United Kingdom

Abstract

This paper introduces a parameterisable, application and platform-independent, hybrid memory sub-system for custom hardware. This memory sub-system consists of a scratchpad memory (SPM) and a custom parallel cache, which exploits data re-use effectively in spite of data dependence. The cache is capable of exploiting spatial locality of memory accesses in two dimensions, making it ideal for video applications. Further, we conducted a case study involving the Quad-tree Structured Pulse Code Modulation (QSDPCM) algorithm, commonly used in MPEG applications. Specifically, the data dependent nature of memory accesses is demonstrated. Using the memory sub-system, performance improvements of up to 1.7x and 1.4x are obtained when the application is implemented on an Altera Stratix 2 chip and a Xilinx Virtex 2 chip respectively, compared to a SPM implementation. In addition, memory savings of up to 3.2x are achieved. These results emphasize the importance of developing dynamic memory sub-systems for custom hardware applications.

1 Introduction

Many video applications require substantial storage and bandwidth to maintain an acceptable quality of service. However, these resources are often limited, particularly on hand-held devices. Further, the existence of multiple video communication and processing standards implies a need for a flexible implementation platform. These factors motivate the acceleration of video compression algorithms, such as the Quad-tree Structured Pulse Code Modulation (QSDPCM) algorithm [4], on FPGA platforms. To accelerate video applications, memory sub-systems are frequently built on top of video frame memories to exploit data re-use.

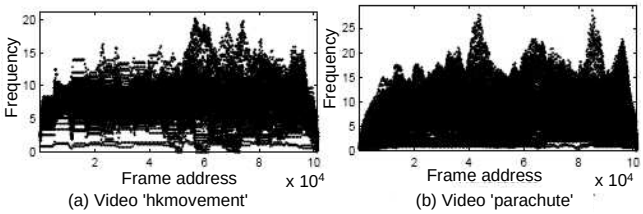

Figure 1: QSDPCM data re-use profiles for two videos.

Indeed, much work [1] has been done for this application on processor-based platforms. These techniques either involve a caching methodology or a fully static memory sub-system.

Our contributions are as follows. First, we introduce a parameterisable, hybrid memory sub-system, consisting of a scratchpad memory (SPM) and a custom parallel cache, which efficiently exploits data re-use in spite of data dependence. Second, the data dependent nature of memory accesses in the QSDPCM application is demonstrated using real video data. Third, compared with an implementation that only employs an SPM, the proposed memory sub-system is found to provide speed-ups of up to 1.7x and 1.4x respectively on two popular FPGA platforms. Further, memory reductions of up to 3.2x are achieved as well.

2 Data dependent memory accesses

In the QSDPCM application, compression is achieved by exploiting spatial correlation in video images. Indeed, previous studies [4] indicate that correlation within image blocks may be more effectively captured by accounting for inter-frame movements, implying that pixel accesses are dependent on the motion vectors of the video sequence. This data dependent feature is demonstrated in Figure 1 for two different video sequences. To handle data depen-

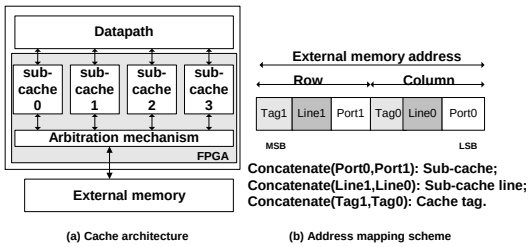

Figure 2: Cache architecture and address mapping scheme.

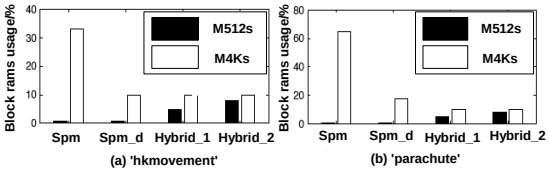

Figure 3: Resource usage results.

dent memory accesses, a custom parallel caching methodology, shown in Figure 2, is employed. This scheme allows parallel accesses to sub-cache banks and arbitrates external memory accesses when multiple sub-cache misses occur. Also, the address mapping scheme exploits two dimensional spatial locality by ensuring that neighbouring pixels map to seperate sub-caches banks.

3 Results

Experiments are carried out to measure the benefits of the hybrid memory sub-system. To make meaningful comparisons, the search window size for each video sequence is set to capture 75% of the motion vectors. The platforms used are the Celoxica RC250 [2] and RC300 [3] boards, which contain the Stratix 2 and Xilinx Virtex 2 FPGAs respectively. Four different schemes are considered. They include *Spm*, which contains an SPM that accomodates the largest possible motion vector, such that data accesses are strictly limited to the SPM. Conversely, *Spm_d* makes use of an SPM that allows accesses to external memory if the required data is not in the SPM. Finally, *Hybrid_1* and *Hybrid_2* are schemes which use SPMs and caches with capacities of 512 and 1024 pixels respectively.

For the hybrid schemes, an initial reduction in run time occurs as the SPM size increases. This result is expected because pixels in the neighbourhood of the reference block, with high re-use frequencies, are buffered. However, an upturn in execution time occurs as SPM size continues to increase because of the overhead incurred to buffer pixels farther away from the reference block. Table 1 shows the realised speed-ups, compared with *Spm_d*. For small inter-

Table 1: Application speed-ups for different image sequences and architectures; video sequences are arranged in order of increasing inter-frame movements.

Video sequence	Window size/ pixels	*Altera Stratix 2*	
		Hybrid_1	*Hybrid_2*
'hkmovement	38	1.068x	1.041x
'matrix'	55	1.358x	1.355x
'starwars'	67	1.427x	1.437x
'parachute'	97	1.637x	1.702x
Image sequence	Window size	*Xilinx Virtex 2*	
		Hybrid_1	*Hybrid_2*
'hkmovement'	38	0.974x	0.977x
'matrix'	55	1.224x	1.275x
'starwars'	67	1.282x	1.344x
'parachute'	97	1.298x	1.387x

frame movements ('hkmovment'), the hybrid schemes use slightly more memory than *Spm_d*, as shown in Figure 3. Conversely, for large inter-frame movements ('parachute'), the hybrid schemes uses up to 3.2x less block RAMs.

4 Conclusion

In this work, performance gains and memory resource savings of up to 1.7x and 3.2x respectively are realised through the adoption of a hybrid memory sub-system on contemporary FPGA platforms. Greater gains are realised for video sequences with greater inter-frame movements. The results emphasise the need for dynamic memory sub-systems in custom hardware to deal with data dependent memory accesses. In future, we intend to develop systematic approaches to determine the optimum parameters of the hybrid memory sub-system.

References

[1] M. Absar and F. Catthoor. Compiler-based approach for exploiting scratch-pad in presence of irregular array access. In *Proceedings of the Design, Automation and Test in Europe*, pages 1162–1167, 2005.

[2] Celoxica, RC250 board specifications, http://www.celoxica.com/products/rc250/default.asp, accessed 11 Jan 2007.

[3] Celoxica, RC300 board specifications, http://www.celoxica.com/techlib/files/cel-w040216143f-257.pdf, accessed 11 Jan 2007.

[4] P. Strobach, D. Schutt, and W. Tengler. Space-variant regular decomposition quadtrees in adaptiveinterframe coding. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 1096–1099, 1988.