Power Tradeoffs in Asynchronous Interfaces

D. K. Arvind*and Kristian Hildingsson
Institute for Computing Systems Architecture
Division of Informatics, The University of Edinburgh
Edinburgh EH9 3JZ, Scotland

Abstract

This paper investigates the influence on the power consumption of partitioning a system-on-a-chip into locally-clocked modules called Temporal Regions, and replacing its global clock with a four-phased asynchronous protocol. Energy dissipation measurements are presented for an implementation of a JPEG encoder chip implemented in 0.35μm CMOS process in a design environment based on commercial EDA tools.

1 Introduction

This paper explores the efficacy of partitioning a system-on-a-chip into Temporal Regions (TR), where the timing within each TR is governed by locally-generated clocks, and the operations between the TRs are controlled by four-phase asynchronous protocols. The removal of the global clock offers some potential advantages: a prominent source of energy consumption is now eliminated; the clocking in each TR can be optimised by distributing the clock over a smaller area and its frequency can be tailored to local requirements in each TR. We first examine the design of one such asynchronous interface, and measure the energy dissipation in the interface and the clock net for different-sized TRs. These results were derived for a JPEG encoder chip (approximately 100,000 gates) implemented in a 0.35μm CMOS process in a design environment based on commercial EDA tools, such as Design Compiler (synthesis) and PowerMill (power measurement) from Synopsys and Silicon Ensemble (place and route) from Cadence Design Systems.

2 The Asynchronous Synchronisation Mechanism

Independently-timed blocks can be synchronised in different ways [Sei94]. A pausible clocking scheme (PCC) [Yun99] is one which is guaranteed to be free of synchronisation failures. Figure 1 illustrates such a scheme for two-way communication, via FIFO buffers, between Temporal Regions, SB1 and SB2. Each synchronous block has two inputs, `R1p` and `A1d` (in the case of SB1), which are synchronised to the local clock in the PCC unit, as shown in Figure 2. The incoming request signals, `Rp` and `Ad`, are synchronised to the FSMs. The synchronisation strategy for the PCC ensures that, under certain timing requirements, there will be no violations of set-up and hold-times. The AFSM performs a two-to-four phase conversion of the handshaking signals. This ensures that both edges of the requesting signals are synchronised to the local clock. A two-way arbiter selects one of the request signals for forwarding to the Mutual Exclusion (ME) block, and the AFSMs are acknowledged with grant signals `G1` and `G2`. The two inputs to the ME block are the request signal, which is forwarded from the arbiter, and the local clock signal. The signal `G` is asserted immediately once the clock signal is low, and the request signal goes high. Once the transition in `G'` is recognised by the arbiter, then it sets the corresponding output grant signal (`G1` or `G2`), and lowers the output request signal `R`, which in turn renders `G` low. In summary, all transitions on the input signal `Rp` to the AFSM results in a request to the arbiter, which is forwarded to the ME and synchronised to the local clock. The AFSM, therefore, performs a two-to-four phase conversion of the handshaking signals.

*Contact person: dka@dcs.ed.ac.uk

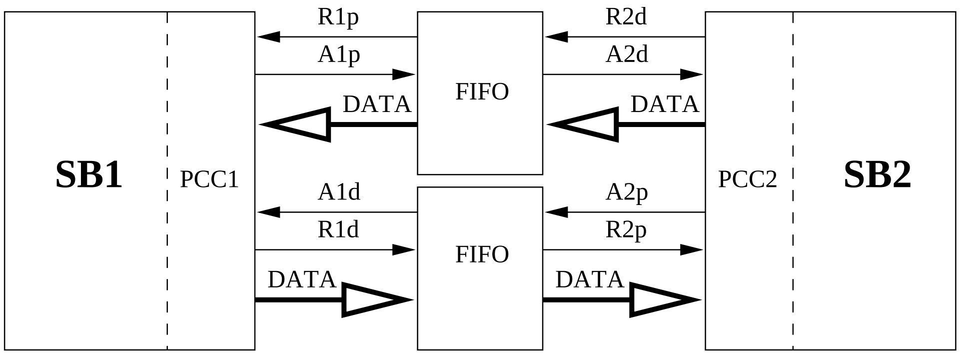

Figure 1: Pausible Clocking Control Scheme (PCC)

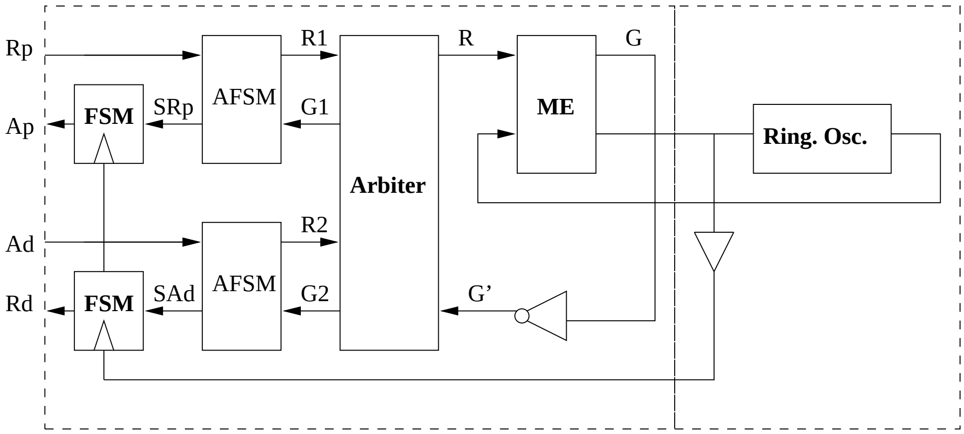

Figure 2: PCC at the Block-level

3 Experiments in Energy Dissipation

We next examine the impact on the energy dissipation of replacing the global clock with a combination of locally-generated ones and asynchronous protocols. This exercise was targetted at a JPEG encoder chip [Kov95] designed in 0.35μm CMOS technology (see Figure 4), using the design environment illustrated in Figure 3. One of the issues of interest is the impact of the different sizes of TRs on the energy dissipation, i.e. the trade-off between the number of partitions and the overhead of the asynchronous protocol. The modules in the JPEG form a linear pipeline, and aggregating numbers of adjacent blocks gives different partitioning scenarios, as shown in Figure 7.

All the modules were separately synthesised using the same timing parameters. In the globally synchronous case (Case A) all the modules were imported into Silicon Ensemble (SE) and the clock tree was routed using CTGen. For a partition of two modules (Case B), modules 1 to 4, and 5 to 11, were imported separately into SE, and CTGen was used to route the clock exclusively within the two partitions. This exercise was repeated for all the eleven partitions.

The clock buffers were extracted in each case and the energy dissipation measured using PowerMill. These measurements were repeated for an implementation of the ring oscillators and the asynchronous protocol, e.g. the overhead for scenario C is due to three ring oscillators (one per partition) and two asynchronous protocols. Figures 8 and 9 plot, for different partitoning scenarios, the energy dissipation in the clock nets, and the asynchronous protocol and ring oscillator, respectively. These two figures are combined in a single graph (Figure 10) which

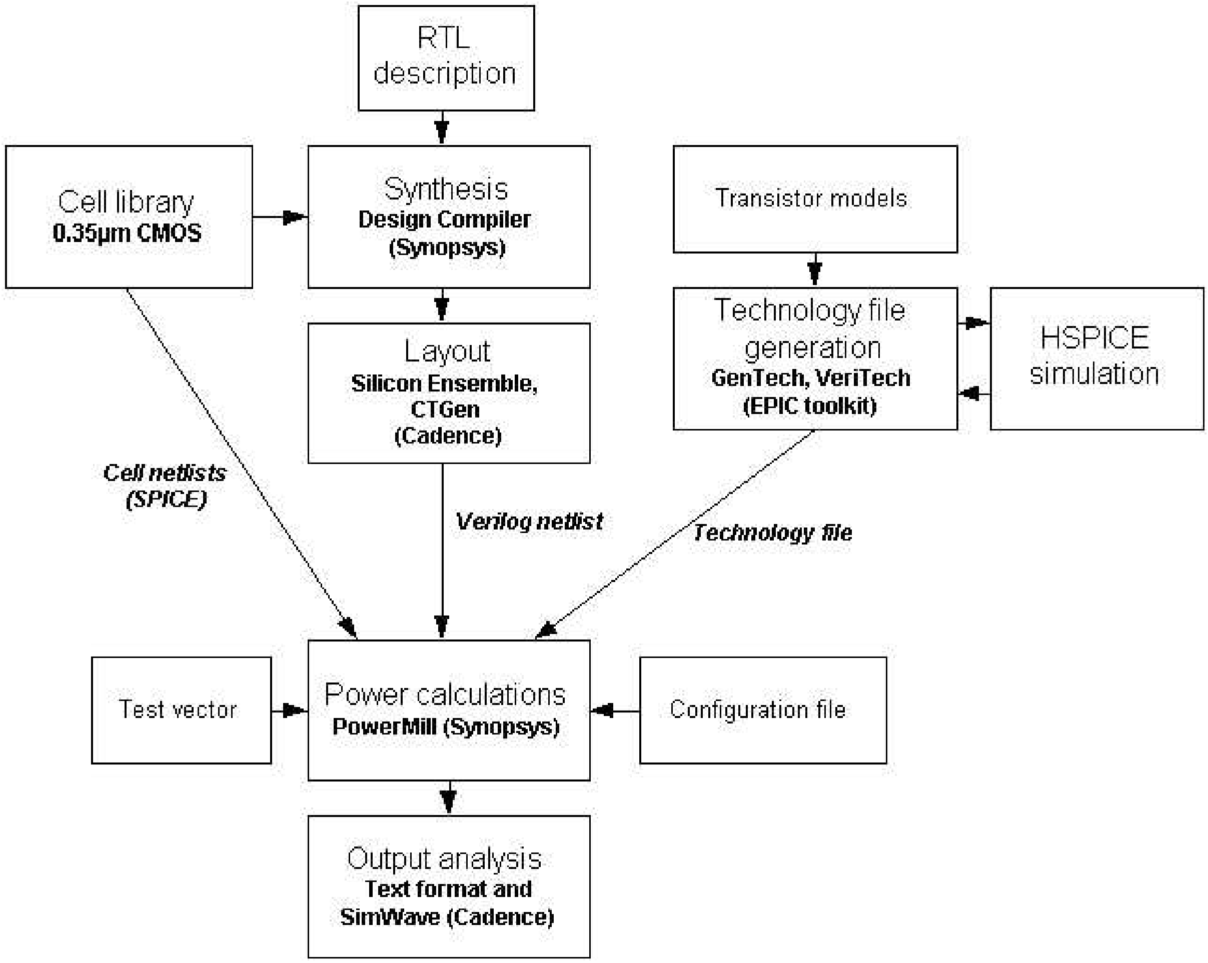

Figure 3: Design Environment

gives the overall energy dissipation.

4 Results

The clock trees were generated automatically using CTGen based on the following timing parameters: clock frequency - 40 MHz, maximum insertion delay - 1.2ns, minimum insertion delay - 0.2ns, maximum clock skew - 0.5ns. The size of the clock tree is sensitive to these timing constraints; for instance, decreasing the maximum insertion delay to 1.0ns results in an increase in power consumption from 1.5mA (RMS) to 5.6mA (RMS) for the DCT block.

The block size and number of clocked elements impacts on the number and sizes of the buffers in its clock tree. Figure 6 illustrates the energy dissipation in the clock trees for different block sizes. The same-sized clock tree is chosen for a range of block sizes; for instance, blocks 12 to 18 require approximately the same clock tree (and therefore the same energy dissipation), although the block sizes range from 12,000 to 26,000 gates. Therefore, halving the DCT blocks (16 and 17) is unlikely to impact the size of clock tree in any major way, and this was confirmed experimentally. This effect also contributes to the increase in energy dissipation when we move from one to two partitions (scenario B) in Figure 8. The individual sizes of the partitions are still large enough

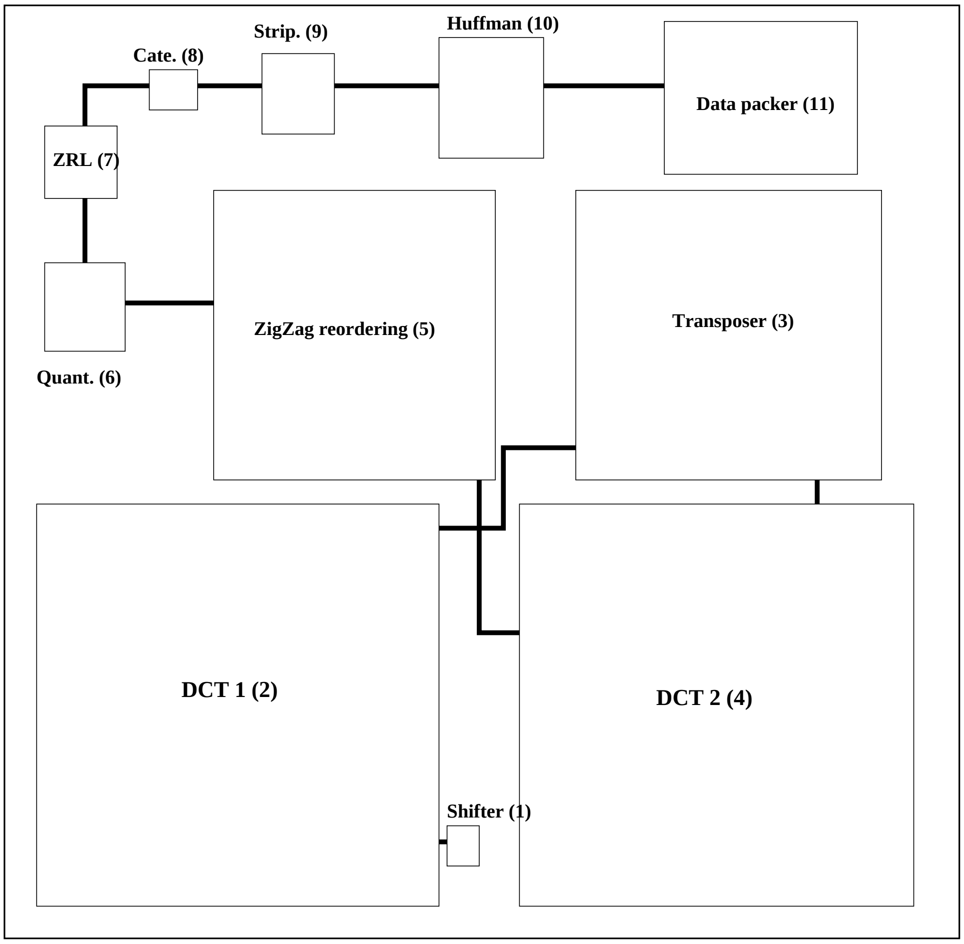

Figure 4: Block diagram of the JPEG encoder

to merit the same size and numbers of clock buffers in each partition. This effectively doubles the number of buffers, and together with the additional synchronisation overhead (which is small in comparison), it results in an increase in energy dissipation for scenario B. The situation improves as the sizes of the partitions decrease, which is accompanied by the choice of smaller and less numerous buffers. Scenario E represents the optimal number of partitions - more than this number results in an increase in the synchronisation overhead without any accompanying decrease in the sizes and numbers of clock buffers.

The FIFO in Figure 1 buffers communications between the Temporal Regions when they operate at different frequencies. The power penalty of the buffers depends on its size, the resolution of the data buses connecting the TRs, and the switching probability of the data communicated over the buses. The globally synchronous JPEG communicates one data item per clock cycle, and the asynchronous protocol will degrade the pipeline throughput. This performance penalty depends on the latency due to the handshaking signals and the phase shift between the local clocks in adjacent TRs. In the worst case, the data will take two clock cycles to communicate if the two clocks operate in opposite phases.

Module No.	Function	Gate Count
1	Shifter	155
2	DCT1	25206
3	Transposer	12285
4	DCT2	25206
5	Zigzag re-ordering	12285
6	Quantiser	1063
7	ZRL encoder	1003
8	Category select	433
9	Strip logic	943
10	Huffman	1404
11	Data packer	7583

Figure 5: Sizes of the blocks in the JPEG encoder

5 Conclusions

This paper has reported the first detailed experiments on the impact of replacing a global clock with Temporal Regions which are co-ordinated by asynchronous protocols. The experiments were conducted on a non-trivial design in a design environment based on commercial EDA tools. The results reflect the influence of the sizes of the Temporal Regions on the energy dissipation in the clock nets and synchronisation overheads, but qualified by the practical limitations of clock tree generators.

Acknowledgements

The authors would like to acknowledge the Institute for Computing Systems Architecture for their support, and the SHEFC-funded RDG grant entitled *Design Clusters for System Level Integration* and Project ALBA for equipment and software grants. Kristian Hildingson would also like to acknowledge a travel grant from the EPSRC-funded Engineering Doctorate programme, and Hitachi Micro Systems Europe for their kind support.

References

[Kov95] Kovac, M. and Ranganathan, N. 1995. “A fully pipelined VLSI architecture for JPEG image compression standard”, *Proc. of the IEEE*, 83(2), Feb. 1995.

[Yun99] Yun, K. Y. and Dooply, A. E. 1999. “Pausible clocking based heterogeneous systems”, In *IEEE Trans. on VLSI Systems*, Vol. 7, Dec. 1999.

[Sei94] Seizovic, J. N. 1994. “Pipeline Synchronisation”, In *Proc. Int. Symp. on Advanced Research in Asynchronous Circuits and Systems*, Nov. 1994.

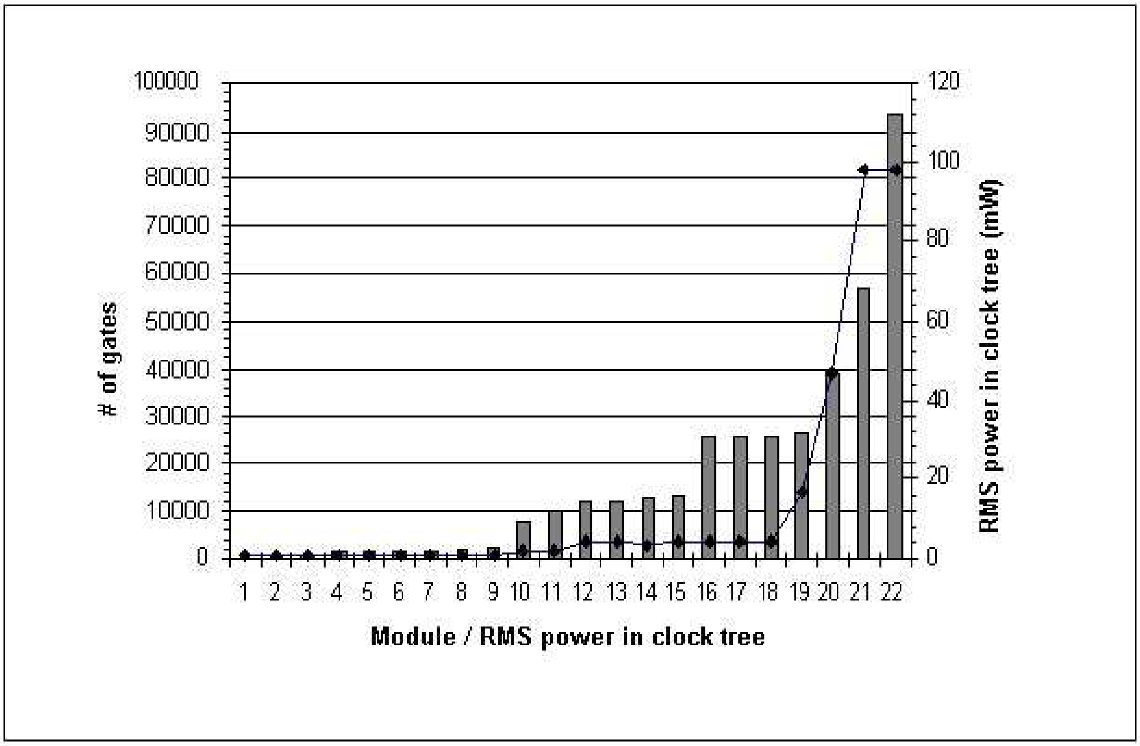

Figure 6: Gate count versus energy dissipation

Case	**Partitioning**
A	1-11 (globally synchronous)
B	1-4, 5-11
C	1-3, 4, 5-11
D	1-2, 3, 4, 5-11
E	1-2, 3, 4, 5-6, 7-11
F	1-2, 3, 4, 5, 6-8, 9-11
G	1-2, 3, 4, 5, 6, 7-8, 9-10, 11
H	1-2, 3, 4, 5, 6, 7, 8, 9-10, 11
I	1-2, 3, 4, 5, 6, 7, 8, 9, 10, 11
J	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

Figure 7: The partitioning scenarios

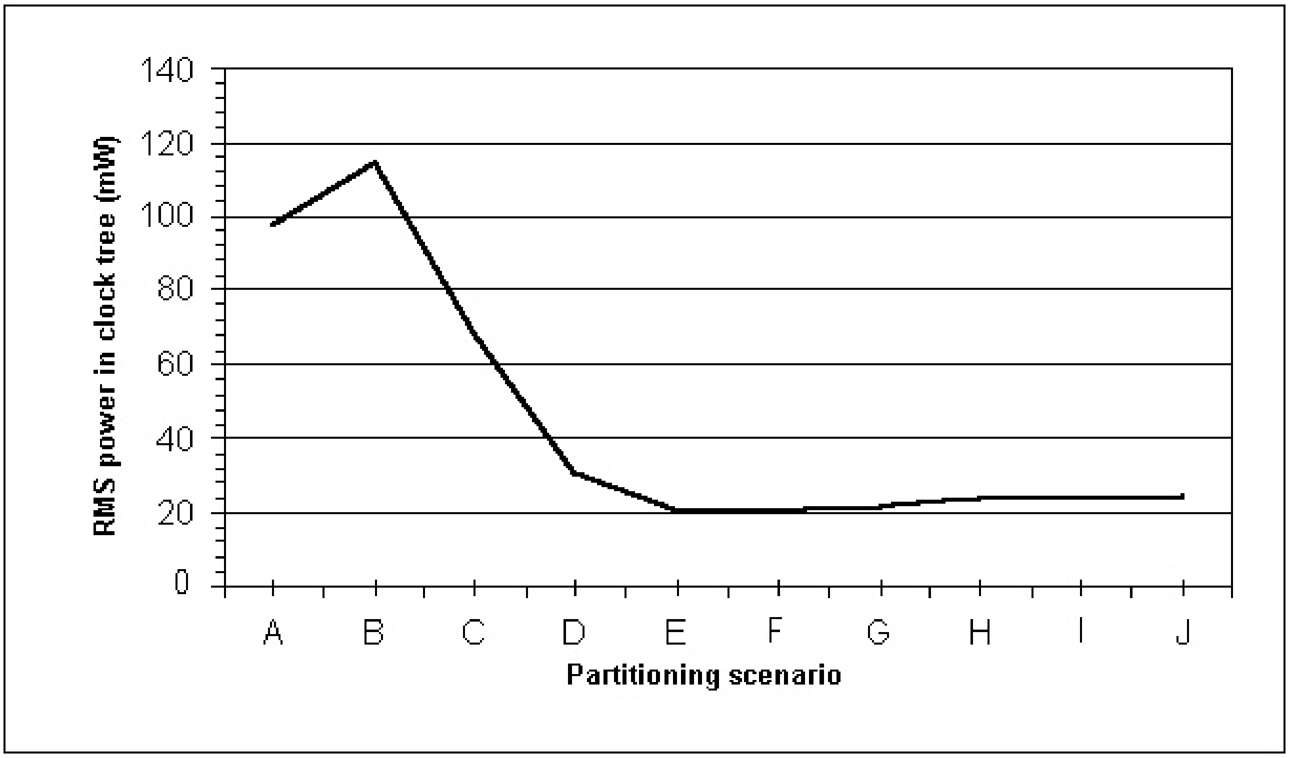

Figure 8: Energy consumption in the clock net

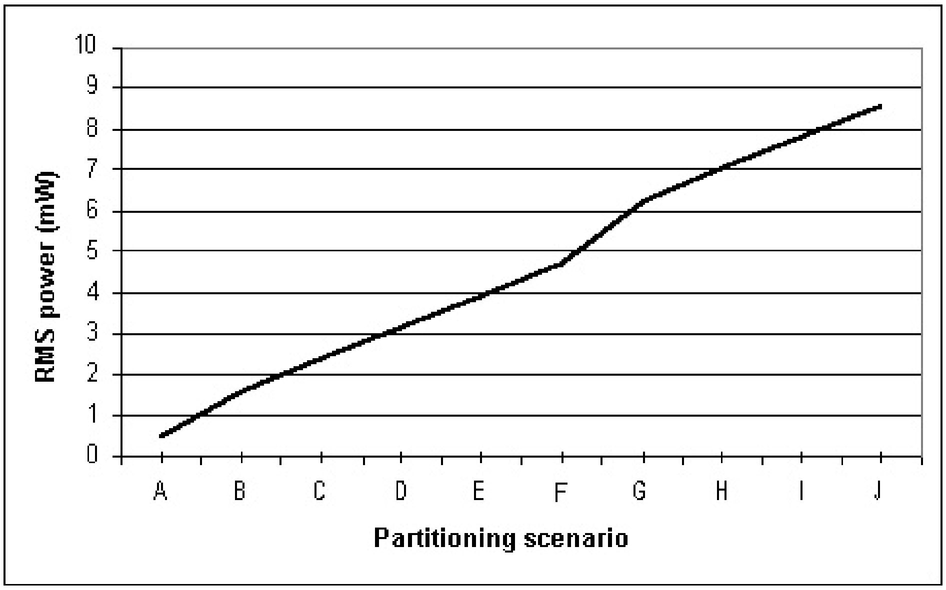

Figure 9: Energy consumption in the asynchronous interface

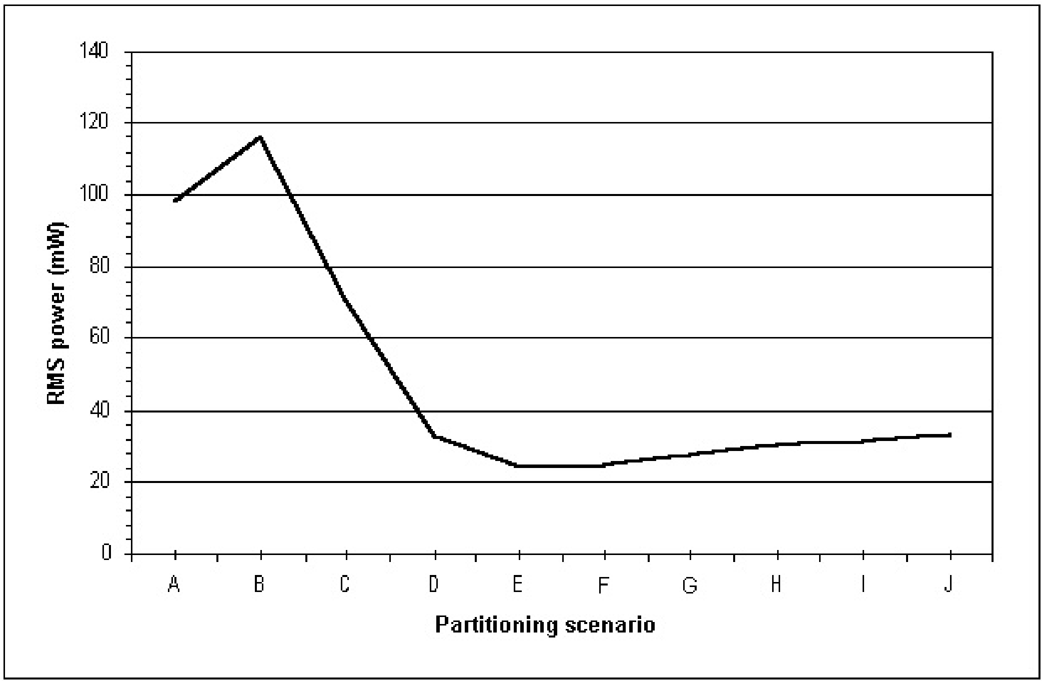

Figure 10: Total energy consumption in clock net and asynchronous interface