On the Impact of Area I/O on Partitioning: A new Perspective

Etienne Hirt, Jean-Pierre Wyss, Andreas Thiel, Daniel Ammann, Michael Scheffler
Claus Habiger, Gerhard Tröster
Electronics Laboratory, ETH Zürich
CH-8092 Zürich, Switzerland
e-mail: hirt@ife.ee.ethz.ch

Abstract

Today's design of computer systems is mainly limited by the achievable I/O bandwidth. Chip designers try to avoid this barrier by designing larger and larger chips. Package designers on the other hand are facing ICs with smaller and smaller pad pitches for more and more I/Os. This traditional separation between chip and package design blocks new solutions. But a close co-operation of chip and package designers allows new partitioning options by combining area I/O with new chip design. The co-operation brings up more and faster I/Os which are easy to connect. The achievable improvements are exemplified on a generic microprocessor system design.

1 Motivation

State of the art processors show ever increasing internal clock rates to improve performance. But the external clock rate as well the I/O bus width hardly match this trend as shown in the roadmap (table 1). It was tried to overcome this discrepancy of off-chip bandwidth to on-chip speed by adding several levels of caches, which enlarges the maximum latency more and more [3]. The larger I/O buses proposed in the roadmap are difficult to connect to the outside. Furthermore, the package parasitics slow the speed down. Therefore, the IC designers as well as the package designers have to co-operate to provide new solutions based on area I/O as proposed in section 2. To illustrate the achievable improvements, they are discussed on a microprocessor system design in section 3.

Year	1998 - 2000	2001 - 2003	2004 - 2006	2007 - 2009	2010 - 2012
On-Chip Frequency [MHz]	200	300	400	500	625
Off-Chip Frequency [MHz]	66	100	100	125	150
I/O Bus width	64	128	128	256	256

Table 1: Roadmap for Cost/Performance Semiconductors[4]

2 Technology

Standard IC configuration with peripheral pads as shown in figure 1 is widely used in industry. The technology is matured with good availability. Nevertheless, it has some significant disadvantages: It shows a low aspect ratio for the core to I/O area, the number of I/Os is relatively low even at small pitches and the speed remains moderate due to mutual inductance as well as other parasitics. These aspects cause a low throughput from chip to chip. Therefore, system designers tend to keep

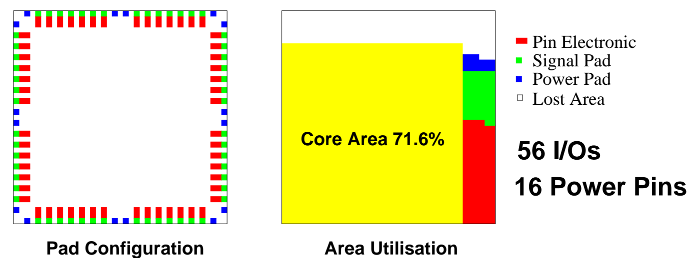

Figure 1: Standard IC Configuration with peripheral pads

all high speed parts on the same chip which leads to large ICs. New concepts as described below allow higher speed systems at moderate cost. They are illustrated using always the same generic chip size and changing the generic I/O placement. This results in a larger core area and therefore less lost area.

2.1 Adaptions to chip design

The core area is improved by placing the pin electronic right under the pads, which is already an aim of I/O buffer designers.
Another approach already used is re-routing standard peripheral ICs to area I/O ICs. Re-routing normally decreases the routing pitch on the next interconnection level. But the number of I/Os remain the same and their capacitive loads are enlarged through longer interconnection distances. And rerouting is only applicable with Flip Chip (FC) mounting.
FC itself allows higher signal speed and reduces ground bounce as well as the chip footprint. But it is difficult to implement for small pad pitches. Therefore, new interconnection concepts using FC attachment have to be considered, which needs co-operation between chip and package designers.

2.2 Alternative I/O configurations

The simplest modification in chip I/O design is to leave the signal pads peripheral, but to spread the power contacts all over the core area as shown in figure 2. The number of I/Os is only slightly increased as the peripheral power pads can be replaced by I/O pads, but ground bounce is reduced.
If the major concern is to increase the number of I/Os, the pin electronic can be left peripheral, but the pads are spread all over the chips as shown in figure 3. The number of I/Os can be increased significantly compared to a traditional chip at a much larger pad pitch. The signal speed may degrade due to longer capacitive loaded interconnections, but the pin electronic remains peripheral. Therefore the ESD protection can be implemented without performance loss and without influencing the core design. If the power pads are also spread to area interconnect, the switching noise is reduced.

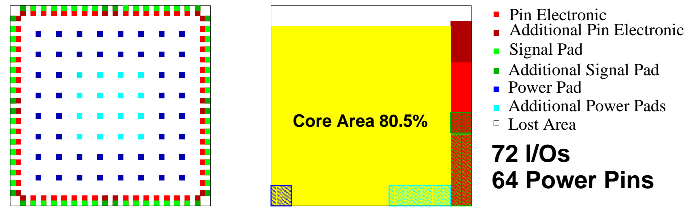

Figure 2: Power distribution on Core

From a system perspective however, the best strategy is to place all signal as well as the power pads in an area arrangement with the pin electronic near the pads. As shown in figure 4, the I/O pads are directly placed where they are connected to the core, reducing the on-chip routing as well as the capacitive loads. Due to the distributed power pads, a good power supply is warranted.

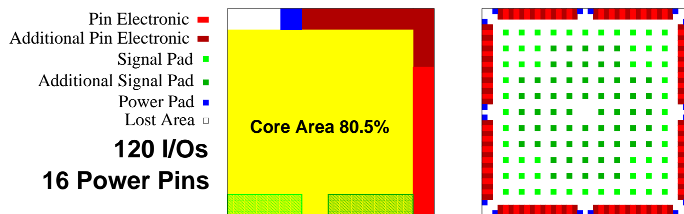

Figure 3: Area IO with peripheral pin electronic

Furthermore, the core area is enlarged and even more I/Os and power pins can be placed. But this concept is completely different from the one followed for years in chip design. Additionally, it splits the core for ESD protection, which is not familiar to chip designers. Nevertheless, it is worth considering these approaches, because they allow new partitioning options as described in section 3.

Figure 4: Area IO with pin electronic near the pads

3 Design study - research vehicle

Today's typical processor systems as shown in figure 5 consists of a CPU with on-chip first level cache running at maximum speed of 300 MHz (f_{CPU}). The host bus is eight bytes wide, running at maximum 100 MHz. It connects the 512kB large second level cache as well as the DRAM interface and the peripheral bus. The system is speed limited through the low number of I/Os at moderate speed. The bandwidth is further degraded because all transfers use the host bus. To

improve the average bandwidth, chip designers tend to increase the first level cache and to add the second level cache on chip. A third level cache would even enlarge the control overhead. This results in big CPU chips and long latency in case of cache miss. Furthermore, the host bus has rather large capacitive load that doesn't allow to increase its speed. Additionally, it deteriorates ground bounce. In spite of the rather slow throughput, the system has the big advantage of having a rather low pincount. Based on the assumptions in table 2 some figures can be calculated using

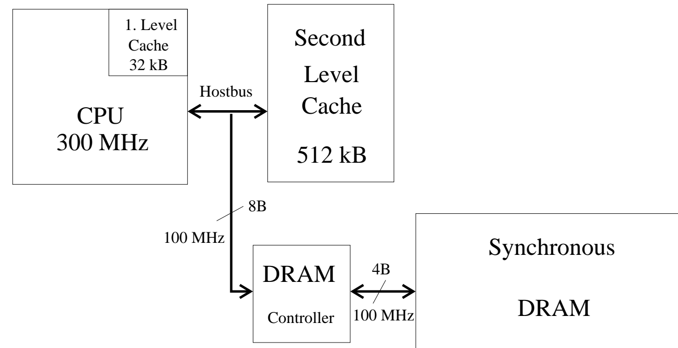

Figure 5: Actual System Design

the formulas (1) - (5), which allows a rough estimation of the required performance. The results shown in table 3 illustrate the bandwidth limitation of the system. By comparing the available bandwidth to the needed throughput it can be seen that preload of data is not possible.

Bytes/command	B_C	2
CPU cycles/command	C_C	3
Data bytes/command	B_D	4
CPU cycles for fi rst level miss	C_f	1
CPU cycles for second level miss	C_s	6
CPU cycles for DRAM addressing	C_D	18+3
DRAM cycle time	T_A	10 ns
DRAM bus width [Bytes]	D_{Bw}	4
Host bus width (Data, Parity, Enable) [Bytes]	H_{Bw}	8
Host address	H_A	29
General purpose pins	GP	60
Core power pins	C_p	74
Signal to Power Ratio	SPR	4

Table 2: Assumptions

$$Maximum\ Latency = C_f + C_s + C_D \quad (1)$$

$$Maximum\ Troughput = D_{Bw}/T_A \quad (2)$$

$$Needed\ Troughput = \frac{f_{CPU}}{C_C} * (B_C + B_D) \quad (3)$$

$$Signal\ Pins = (H_{Bw} * 10) + H_A + GP \quad (4)$$

$$Power\ Pins = \left(\frac{Signal\ Pins}{SPR}\right) * 2 + C_p \quad (5)$$

Maximum Latency	28 CPU cycles
Maximum Troughput	400 MB/s
Needed Troughput	600 MB/s
Signal Pins	169
Power Pins	158

Table 3: Speed, performance and I/O requirements of today's system

3.1 Improved system architecture

The system presented above can be improved by means of new partitioning options [1], using the technology proposed in section 2. To explore the possibilities we propose the system shown in figure 6 as a basis for discussions. The first level cache is moved off-chip which improves CPU die yield. The cache is splitted into a data and instruction part and its proposed size of 256 kB is in the range of today's second level cache. But its access time fits to the processor due to the private bus to the CPU. The much smaller TAG RAM remains on chip. The bandwidth to the main memory is improved by factor four by doubling the bus size as well as using two independant banks. This throughput will not be slowed down by accessing peripherals because there is a separate four byte wide bus for this purpose. Furthermore, the DRAM controller is implemented on the CPU chip ensuring short latency. Based on the assumptions in table 2 and 4 and using the formulas (1) - (6), the performance figures are calculated and compared to today's system design in table 5.

$$Signal\ Pins = 2 * (C_{Ctrl} + C_{Bw} * 9) + 2 * (D_{Bw} * 8 + D_{Ctrl}) + (H_{Bw} * 10) + H_A + GP \quad (6)$$

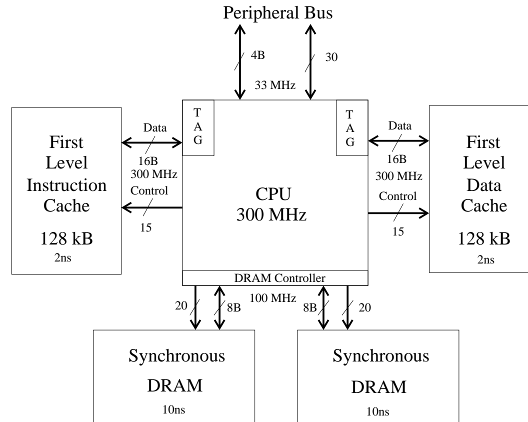

Figure 6: Improved System Design proposition

This comparison shows that the throughput is improved by factor four and the latency by two at the cost of three times the I/Os. Placing this number of I/Os is possible on a common IC with 10 * 10 mm^2 chip size. Considering the pin electronic peripheral, their pitch is 65 μm. This is hardly possible for wirebonding. But the pad pitch in our example with its 344 μm is much larger. Even a μBGA could be used as package.

Cache bus width (Data, Enable) [Bytes]	C_{Bw}	16
Cache control	C_{Ctrl}	15
CPU cycles for DRAM addressing	C_D	18
DRAM bus width [Bytes]	D_{Bw}	8
DRAM control	D_{Ctrl}	20
Host bus width (Data, Parity, Enable) [Bytes]	H_{Bw}	4
Host address	H_A	30
Signal to Power Ratio	SPR	8

Table 4: Assumptions for Improved System

	New System	Improvement
Maximum Latency	19 CPU cycles	32 %
Maximum Troughput	1600 MB/s	300%
Needed Troughput	600 MB/s	0%
Signal Pins	616	-265 %
Power Pins	228	-44%

Table 5: Performance Figures and Comparison

Moving the first level cache off chip improves the CPU die yield. Together with the savings of the second level cache, the higher price of the first level caches should become outbalanced.
However, the increased number of nets on the motherboard as well as their highest speed (300 MHz) makes it more complex and needs concurrent design of IC and package. However, these disadvantages are overcome by building a multi chip module including the CPU and the cache. This module reduces pin count and relieves the high speed design from the OEMs[2].

4 Conclusions

It has been shown that good co-operation between chip and package designers opens up promising new technologies. These technologies enable in turn new partitioning options and therefore better system designs. These opportunities have been illustrated on a processor system where the throughput has been increased by factor four and the latency has been halved. Being not more expensive at much better performance than today's systems, it is an example for a cost-effective high-performance solution.

References

[1] S. Banerjia and al. Issues in partitioning integrated circuits for mcm-d/flip-chip technology. In *IEEE Multi Chip Module Conference*, pages 154–160, 1996.

[2] E. Hirt and al. A pentium based mcm for embedded computing. In *11th European Microelectronics Conference*, pages 516–523. International Society for Hybrid Microelectronics, 1997.

[3] D. Patterson and al. A case for intelligent ram. *IEEE Micro*, pages 34–44, Apr. 1997.

[4] SIA. Sematech roadmap. 1997.