WHITE PAPER

Chronos Power- The Next Generation

By S. Giaconi, Chronos Tech

Introduction:

System scaling, enabled by the continuing advances in semiconductor device fabrication technology, is getting more and more challenged by limited on-die resources such as interconnect bandwidth and power. Latest System-on-Chips (SoCs) require a seamless integration of big-data and instant data considering the growing need of cloud computing within the system. Instant data generation requires ultra-low-power devices with "always-on" features at the same time with high-performance device that can generate the data instantly; while big-data requires abundant computing and memory resources to generate the service and information that clients need.

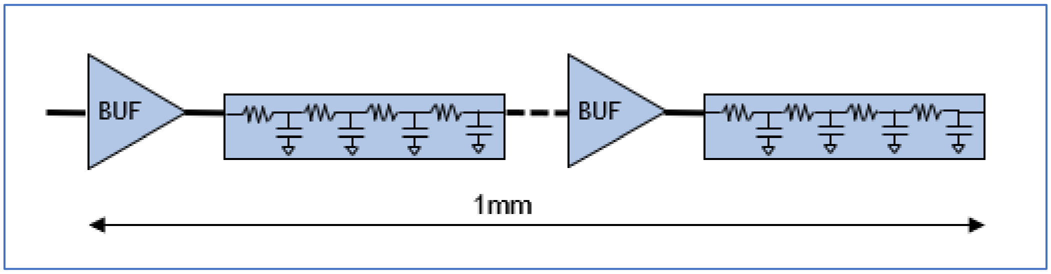
Power constraints apply in different ways for each application:

- **Mobile computing**: more performance and functionality at constant energy (battery limited)
- **High-Performance computing**: more performance at constant power density (thermal limited)
- **Autonomous sensing and computing (Internet-of-Things, IoT)**: reduced leakage and variability

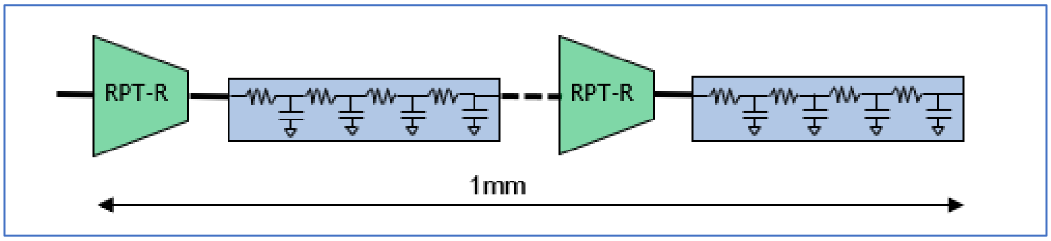
Chronos technology, when applied at architectural level, can enable the next generation SoC, tackling the power challenges affecting the different applications.

Analyzing the power in the interconnect:

To compare power performance, we start analyzing the dynamic power used on a simple connection link. In this example we consider a connection composed by a 32bit bus plus a clock line over 1mm in FinFET technology.

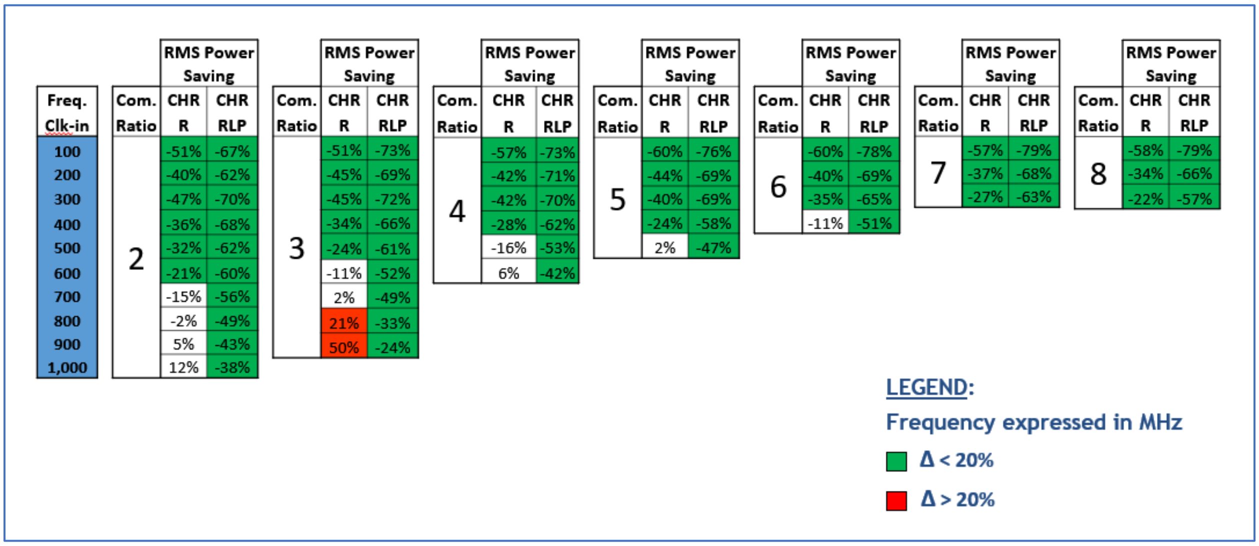


To standardize the comparison between a regular link and a Chronos link we choose the same metal layer for routing (with same metal width and same metal spacing), a clock frequency (at the source) varying from 100MHz to 1GHz, maximum activity factor, maximum crosstalk and number of buffers (or repeaters in the case of Chronos) adjusted to guarantee the maximum transition time required by the specific technology.



Simulations were performed, and measurements were taken with regular link as well as with Chronos Robust link, or Chronos Robust Low-Power Link.

It is important to notice that Chronos links can compress the data up to their individual Maximum Equivalent Speed (MSPD), a parameter dependent on the process technology used and the Chronos library selected (i.e. Chronos Robust vs. Chronos Robust Low-Power). For each input frequency multiple Chronos results will be shown, each one with a different compression ratio. With a higher compression ratio, data will travel on the Chronos link at a faster speed, but use less parallel data signals.



		RMS Power Saving			RMS Power Saving			RMS Power Saving			RMS Power Saving			RMS Power Saving			RMS Power Saving			RMS Power Saving	
Freq. Clk-in	Com. Ratio	CHR R	CHR RLP	Com. Ratio	CHR R	CHR RLP	Com. Ratio	CHR R	CHR RLP	Com. Ratio	CHR R	CHR RLP	Com. Ratio	CHR R	CHR RLP	Com. Ratio	CHR R	CHR RLP	Com. Ratio	CHR R	CHR RLP
100	2	-51%	-67%	3	-51%	-73%	4	-57%	-73%	5	-60%	-76%	6	-60%	-78%	7	-57%	-79%	8	-58%	-79%
200		-40%	-62%		-45%	-69%		-42%	-71%		-44%	-69%		-40%	-69%		-37%	-68%		-34%	-66%
300		-47%	-70%		-45%	-72%		-42%	-70%		-40%	-69%		-35%	-65%		-27%	-63%		-22%	-57%
400		-36%	-68%		-34%	-66%		-28%	-62%		-24%	-58%		-11%	-51%						
500		-32%	-62%		-24%	-61%		-16%	-53%		2%	-47%									
600		-21%	-60%		-11%	-52%		6%	-42%												
700		-15%	-56%		2%	-49%															
800		-2%	-49%		21%	-33%															
900		5%	-43%		50%	-24%															
1,000		12%	-38%																		

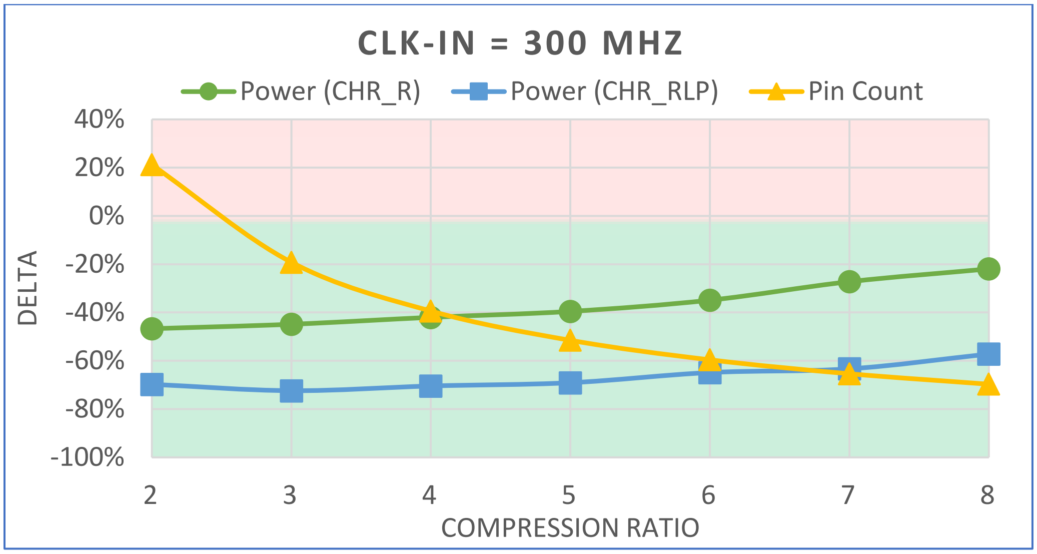
LEGEND:

Frequency expressed in MHz

🟩 Δ < 20%

🟥 Δ > 20%

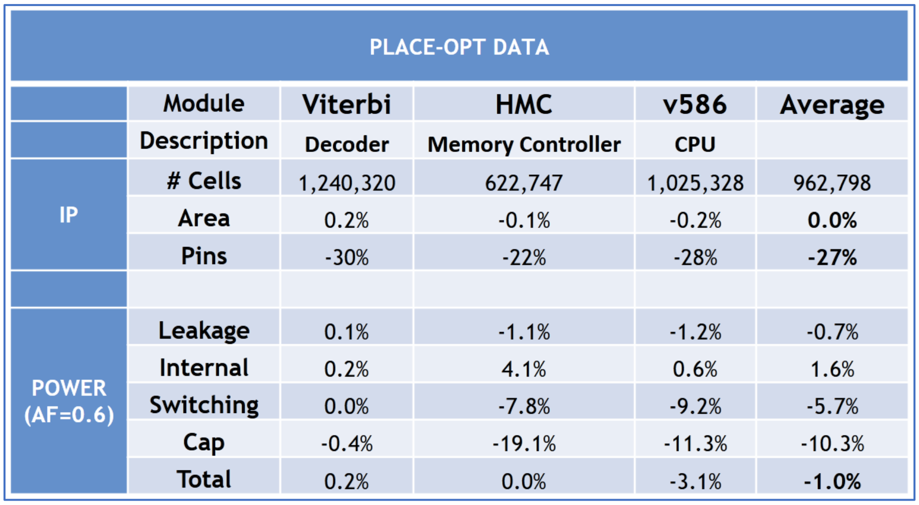
The picture above shows the percentage of difference in RMS power between the Chronos link and the original regular bus link. It is interesting to notice how with Chronos Robust, on average, the RMS dynamic power of the link, described before, is reduced. Chronos Robust Low-Power increases the advantage even further.



The graph above shows the impact of the compression ratio on pin count and power when compared to a regular implementation considering an input clock of 300MHz. The higher the compression ratio, the higher the reduction in pin count, at the expenses of a little increase in power. This flexibility gives the design team the ability to tune the SoC implementation, trading area vs power.

Gasket analysis:

Connecting IPs through Chronos requires the insertion of a gasket, through a process known as *Chronification*, to enable the delay insensitive coding and compression in the communication channel. To establish a fair comparison before and after the *Chronification* process it is good practice to analyze the impact of the gasket itself.



PLACE-OPT DATA

	Module	Viterbi	HMC	v586	Average
	Description	Decoder	Memory Controller	CPU	
IP	# Cells	1,240,320	622,747	1,025,328	962,798
	Area	0.2%	-0.1%	-0.2%	**0.0%**
	Pins	-30%	-22%	-28%	**-27%**
POWER (AF=0.6)	Leakage	0.1%	-1.1%	-1.2%	-0.7%
	Internal	0.2%	4.1%	0.6%	1.6%
	Switching	0.0%	-7.8%	-9.2%	-5.7%
	Cap	-0.4%	-19.1%	-11.3%	-10.3%
	Total	0.2%	0.0%	-3.1%	**-1.0%**

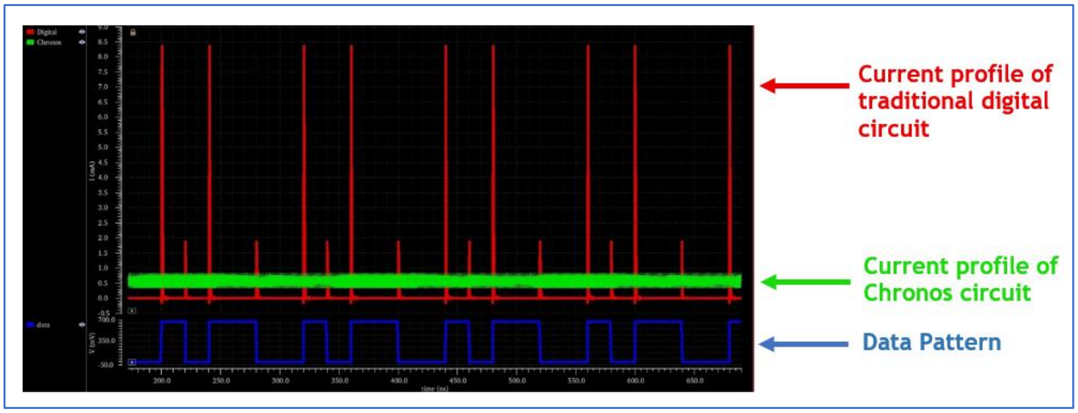
Three different IPs have been considered in this example: A Viterbi decoder (with 1.2M cells), a cube memory controller (HMC, with 600K cells) and a CPU (v586, with 1M cells).

The table above shows the data collected after Place-Optimization step in the flow. The impact of Chronos changes slightly from IP to IP, but we can say that the average area impact of the gaskets is practically negligible. The main justification comes from the fact that even if an extra layer of logic has been added at the boundary of each IP, there is a significant relaxation to internal IP synchronous timing paths associated with the *gasketized* ports, bringing down the IP cell count and area. Power numbers seem also to be very close before and after Chronos insertion (in this example an activity factor of 0.6 has been used to assess switching power within the IPs), reasserting the minimal impact of the gasket itself. One common data factor shown in each IP is the reduction in pin count (because of Chronos Clockless Compression Technology), and a reduction in total capacitance.

Other factors contributing to power:

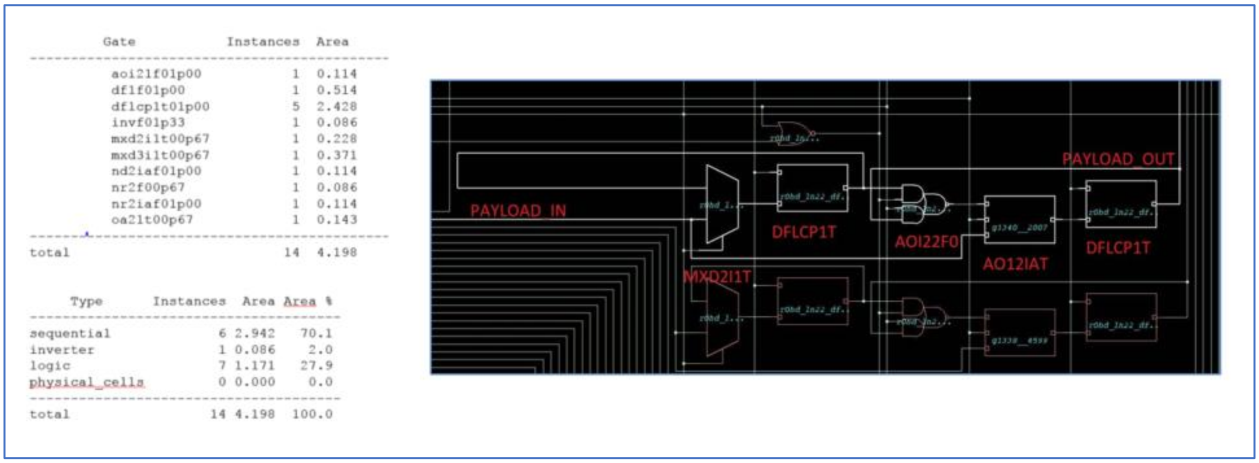
A significant power user in a complex SoC is the clock distribution network. When Chronos links are used to connect IPs, significant reduction in clock distribution is achieved. With Chronos, inter-IP clock lines don't need to be balanced and have no distribution to pipeline flops, reducing significantly the clock tree power required compared to a standard SoC implementation.

A further benefit of Chronos technology is the change in peak power requirement for the interconnect.



In the picture above, a simulation shows the difference between the current profile of a simple standard digital channel, vs. the current profile of the same channel implemented with Chronos. It is interesting to notice how much the peak power is reduced by using Chronos technology, lowering EMI emission and at the same time, relaxing significantly the requirements for the power grid.

Another source of power and area usage in SoC implementation is the insertion of pipeline stages. If timing cannot be met for fast and/or distant IP interfaces, pipelining becomes necessary. In AMBA AXI, for example, Register Slices are inserted to re-time distant interfaces. The picture below shows the complexity of a synthesized implementation of a single bit register slice, composed of 6 sequential elements plus 8 logic elements.



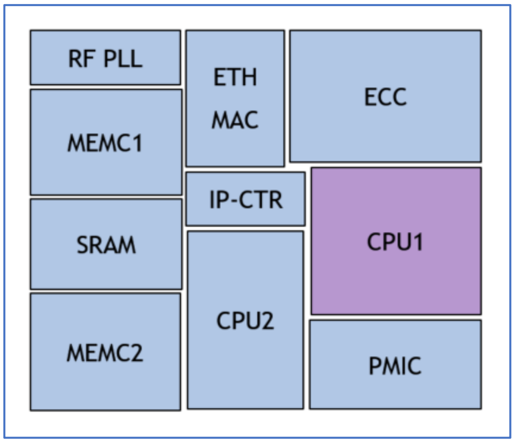
```
        Gate              Instances  Area
-------------------------------------------
          aoi21f01p00             1  0.114
          df1f01p00               1  0.514
          dflcp1t01p00            5  2.428
          invf01p33               1  0.086
          mxd2i1t00p67            1  0.228
          mxd3i1t00p67            1  0.371
          nd2iaf01p00             1  0.114
          nr2f00p67               1  0.086
          nr2iaf01p00             1  0.114
          oa21t00p67              1  0.143
-------------------------------------------
total                            14  4.198


     Type        Instances   Area Area %
-------------------------------------
sequential               6 2.942   70.1
inverter                 1 0.086    2.0
logic                    7 1.171   27.9
physical_cells           0 0.000    0.0
-------------------------------------
total                   14 4.198  100.0
```

Chronos implementation does not need the insertion of Register Slices, Chronos repeaters already act as memory elements behaving like asynchronous FIFOs between the two IPs interfaces, but using a much smaller footprint. Additionally, since IP clocks have no phase relationship, no distribution to top level physical midpoint is needed to maximize clock root common path.

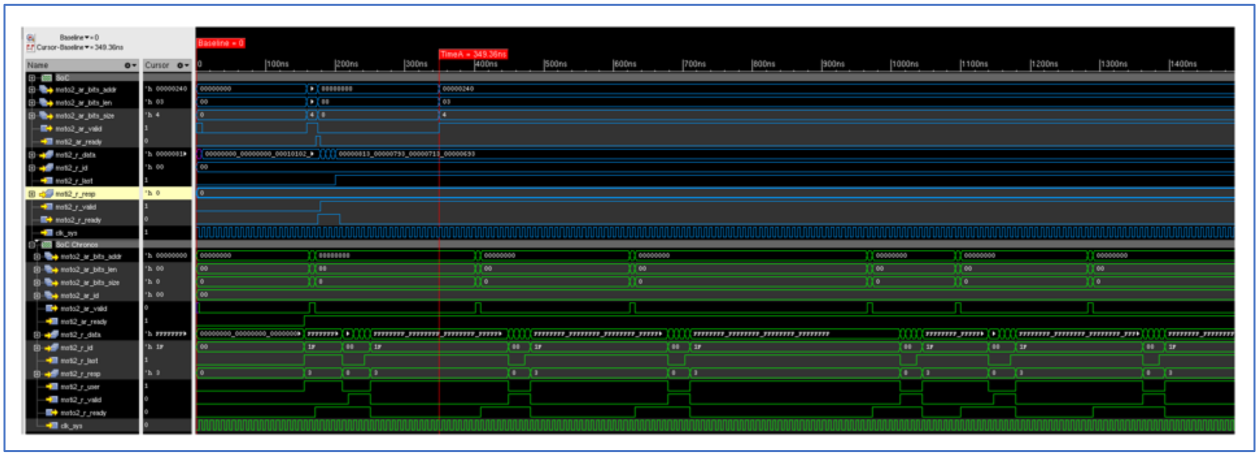
Bringing the savings to the next level:

Delay insensitivity of the Chronos link enables a much bigger potential power saving for a complex SoC, let's consider the following example:

The picture on the right shows the floorplan of a SoC (more specifically, a RISC-V implementation with AMBA AXI busses) where CPU1 is the Rocket-CPU (main brain of the RISC-V system), and IP-CTRL is a tile containing the crossbar to route the data to the different IPs. In the original implementation, CPU1 and IP-CTRL share the same system clock (CLK_{Sys}). In this implementation we use an internal local clock for CPU1 with frequency equal to 0.973 times the frequency of CLK_{Sys}; de facto making the two IPs completely asynchronous to each other.



Then we proceeded to run the same top level SoC simulations for both implementations: with and without Chronos insertion. Below we can see the results for both implementations. The blue waveforms represent the signaling for AR-Master and R-Master AXI interfaces for the original implementation (without Chronos). As we can observe, the AXI transaction stalls because the interfaces cannot handle asynchronous exchange of data.



The green waveforms instead show the same simulation results this time with Chronos inserted. As we can observe, the AXI transaction keeps working smoothly: the integrity of the AXI SoC protocol is maintained and the respective Valid and Ready signals per channel, regulate the throughput, so that there is no data loss.

This experiment unveils the possibility of a much deeper power saving architecture when Chronos technology is used at SoC level. Each IP (or even portion of IP) can independently scale their own voltage and clock frequencies by the usage of local voltage islands and local clock generations. If the architecture is able to measure requirements in different SoC modes for each IP, and it can table performance requirements and schedule them accordingly, the SoC can perform aggressive Adaptive Voltage & Frequency Scaling (AVFS). This will maintain proper functionality while significantly reducing the power consumption. This approach can also be used to easily integrate asynchronous IPs within traditional SoC such as sensor interfaces and/or neuromorphic engines which are asynchronous in nature.

Conclusions:

This paper highlights the impact of Chronos technology to the power challenges in modern SoC design. In many cases Chronos channels can relax the power on the interconnect. Furthermore, if the technology is applied at architectural level, it can dramatically reduce the whole chip power budget for future SoCs, enabling more options to the chip architect and the power management team.