3.5 Godson-3B1500: A 32nm 1.35GHz 40W 172.8GFLOPS 8-Core Processor

Weiwu Hu[1,2], Yifu Zhang[1,2], Liang Yang[2], Baoxia Fan[2], Yunji Chen[1,2], Shiqiang Zhong[2], Huandong Wang[2], Zichu Qi[1,2], Pengyu Wang[1,2], Xiang Gao[2], Xu Yang[2], Bin Xiao[1,2], Hongsheng Wang[2], Zongren Yang[1,2], Liqiong Yang[1,2], Shuai Chen[1,2]

[1]Chinese Academy of Sciences, Beijing, China,
[2]Loongson Technology, Beijing, China

Godson-3B1500 is an 8-core microprocessor product of Loongson Technology™. It is fabricated in 32nm 10 Cu-layer high-κ metal-gate (HKMG) low-power bulk CMOS, and contains 1.14 billion transistors within 182.5mm² die area. Through numerous design improvements, Godson-3B1500 is able to support a wide voltage range from 1.0V to 1.3V, at frequencies ranging from 1.0GHz to 1.5GHz, achieving 172.8GFLOPS at 1.35GHz, with nearly 40W power dissipation. This represents a 35% power-efficiency improvement over the previous design, Godson-3B [1].

Godson-3B1500 maintains a two-node eight-core architecture with an enhanced GS464V core (which is MIPS64-compatible with vector extensions) [2]. However, this design contains two primary changes in its architecture. The first is a modification of the memory hierarchy. The last-level cache (LLC) is increased from 4MB to 8MB, and the introduction of a 4-way 128KB private victim cache in each core reduces data access latency significantly. A low-cost asynchronous FIFO between every core and uncore serves to isolate the core in both the frequency and voltage domains. The second architecture change is enhanced high-speed I/O. The point-to-point HyperTransport (HT) is updated from 1.0 to 2.0, and the memory access interface improves from DDRII 800 to DDRIII 1200, accompanied by a heterogeneous multi-channel controller architecture [3].

The migration from a 65nm process to the 32nm HKMG process provides an additional 14.7% performance boost, but introduces many new challenges in the physical design phase. Structured poly orientation and M1 layer routing are preferred in standard cell design under strict design rules; power, ground and signal wires are enhanced for EM tolerance; considerable floorplan and handcrafted feedthrough placement and routing help to minimize wire delay including crosstalk effects. The number of PVT corners for design verification increased to 17 due to temperature inversion and wire variation, underscoring the importance of multi-corner concurrent timing fixes for design closure. In addition, in contrast to previous implementations, stringent timing constraints are specified for three typical working modes: low-power (1.0V), normal (1.1V) and turbo (1.25V). Hence, multi-point timing optimization enables better scaling behavior across the full operating range.

Furthermore, new voltage detector and processor monitor circuits are integrated to monitor on-chip variation effectively along with thermal sensors. The voltage detector measures the activity of a ring oscillator under test in a configured timing window, and converts the oscillation count to a biased voltage value, which can be observed externally to show short-term or long-term internal IR-drops, as shown in Fig. 3.5.1. The process monitor also contains ring oscillators which consist of unmixed PMOS and NMOS transistors with various thresholds, and is used to check whether the chip remains in the pre-defined process limit during wafer-level test or lifetime debug analysis.

Power dissipation of Godson-3B1500 is another major concern for this 32nm design with nearly double the transistor count of its predecessor. The power consumption breakdown for each component under a typical benchmark is shown in Fig. 3.5.2. Three quarters of the power is attributed to eight processor cores, in which dynamic and static power are split in a 9:5 ratio. This necessitates support for core-level power gating and dynamic voltage scaling to reduce both static and dynamic power. However, unlike other complex power networks [2], a simple yet flexible power and ground system is implemented in Godson-3B1500, as illustrated in Fig. 3.5.3. For die area reduction and IR-drop management, neither power gating cells nor voltage regulators are embedded in the chip. Instead, the VDDCORE power network in each core is isolated and directly supplied by independent VDDCORE bumps. Package-level and board-level co-design can determine the granularity of power management in the number of cores, based on sets of on-board voltage supply. Meanwhile, great design efforts are taken for the communication interface between core and uncore elements, both for timing verification during voltage scaling and state isolation in power gating mode.

The clocking scheme of Godson-3B1500 is shown in Fig. 3.5.4. The central PLL takes an external scattered clock "rclk" (normally at 33MHz) as reference, and generates a global reference clock "gclk" (200MHz) for two next-stage PLLs. The core clock "cclk" (1.5GHz) generated by the CORE PLL propagates to each processor core sequentially, gets reshaped in each dynamic frequency scaling unit, and then distributed to flops, within a worst case of sub-10ps skew by the global mesh and gated local tree. Node clock "nclk" (1.0GHz) from NODE PLL is distributed to the uncore modules by a structured and balanced clock tree, with 14 digital-controlled delay lines (DCDL) inserted at the root of each module for on-chip detection by skew measurement circuitry and fuse-based post-silicon adjustment. The decoupled multi-clock domain scheme enables a frequency boost or drop in each processor core independently, which makes better use of the globally asynchronous locally synchronous (GALS) features in architecture.

The two-stage cascaded clock generation architecture brings more flexibility and efficiency. Bandwidth can be easily optimized for noise rejection. At every stage, a self-biased technique is adopted inside each PLL, which enables a wide input and output frequency feature with low jitter noise. The reference frequency for the cascaded PLL ranges from 1MHz to 100MHz, and the output clock frequency can be up to 3.2GHz. A differential VCO architecture, together with a unique charge pump and switched-capacitor loop filter, also keeps the PLL jitter small. Fig. 3.5.5 shows the schematic and measurement results of the PLL circuitry, in which the output clock at 1.6GHz (with 2MHz reference frequency) has an RMS jitter and peak-to-peak jitter of less than 1.23ps and 12.46ps, respectively.

The HT PHY in Godson-3B1500 achieves a maximum bandwidth of 22.4GB/s, with up to 2.8Gb/pin/s, with BER of less than 10^{-15}. The transmitter adopts a voltage-mode driver and supports 2-tap pre-emphasis and impedance matching to mitigate adverse effects of the channel. Two topologies are used in the receiver for source-synchronous clock and data recovery (CDR): one is the simple direct sampling in low-power mode, and the other is all-digital DLL-based CDR for channel skew compensation in high-speed mode. Fig. 3.5.6 shows the eye diagram of 2.8Gb/s data over a 20cm channel without pre-emphasis and with 3dB pre-emphasis, respectively.

The DDR2/DDR3 combo PHY provides two 64b high-bandwidth memory access interfaces with up to 153.6Gb/s. For various loading conditions of DIMMs, dynamic off-chip driver (OCD) impedance (34-40Ω range with 1Ω step) and on-die termination (ODT) impedance (60-120Ω range with 5Ω step) are supported for termination impedance matching, and the capacitance of IO is minimized with the ODT and the OCD merged together. Dynamic output slew-rate control is also provided for enhanced signal integrity in the same way. Additionally, a self-calibration scheme is included to provide accurate output and termination impedance in different process corners and dynamic environments.

References:
[1] W. Hu, *et al.*, "Godson-3B: A 1GHz 40W 8-Core 128GFLOPS Processor in 65nm CMOS", *ISSCC Dig. Tech. Papers*, pp. 76-78, 2011.
[2] W. Hu and Y. Chen, "GS464V: A High-Performance Low-Power XPU with 512-Bit Vector Extension," *Hot Chips Symposium*, 2010.
[3] G. Zhang, *et al.*, "Heterogeneous Multi-Channel: Fine-Grained DRAM Control for Both System Performance and Power Efficiency," *IEEE/ACM Design Automation Conf.*, pp. 876-881, 2012.
[4] R. Jotwani, *et al.*, "An x86-64 Core in 32 nm SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 162-172, 2011.
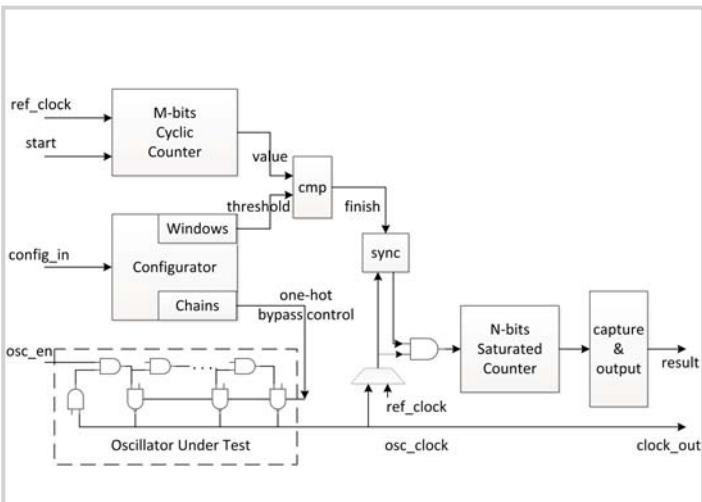
3

Figure 3.5.1: A general schematic of the ring oscillator-based voltage detector.
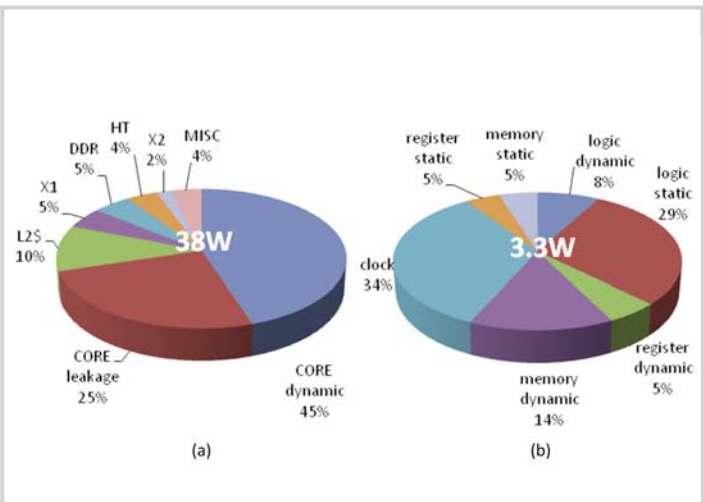

Figure 3.5.2: Godson-3B1500 power consumption breakdown (a) for each component in the whole chip, (b) in the core.
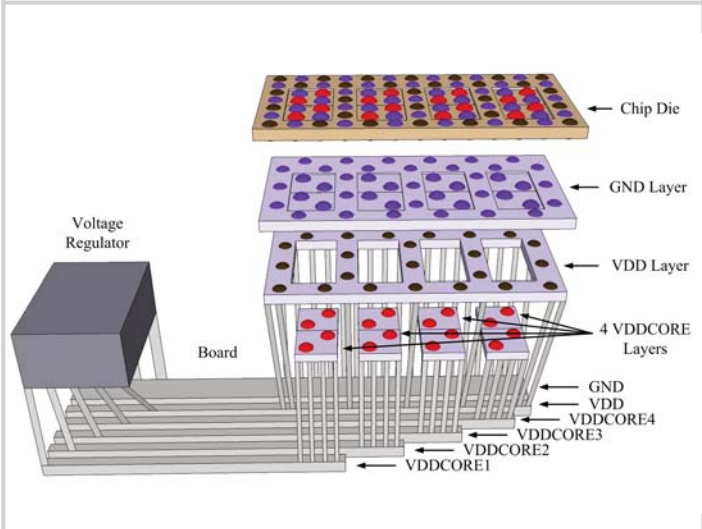

Figure 3.5.3: An illustration of a four-plate VDDCORE power supply system, where each plate is shared by two adjacent cores.
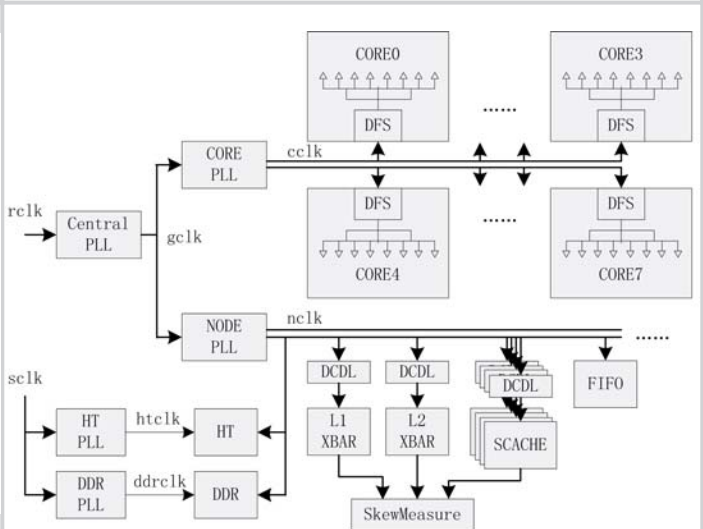

Figure 3.5.4: The clocking scheme of Godson-3B1500.
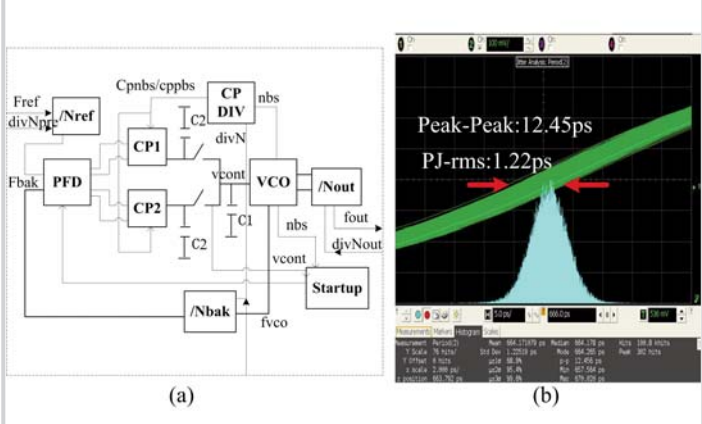

Figure 3.5.5: (a) Schematic and (b) output jitter distribution of the cascaded PLL.
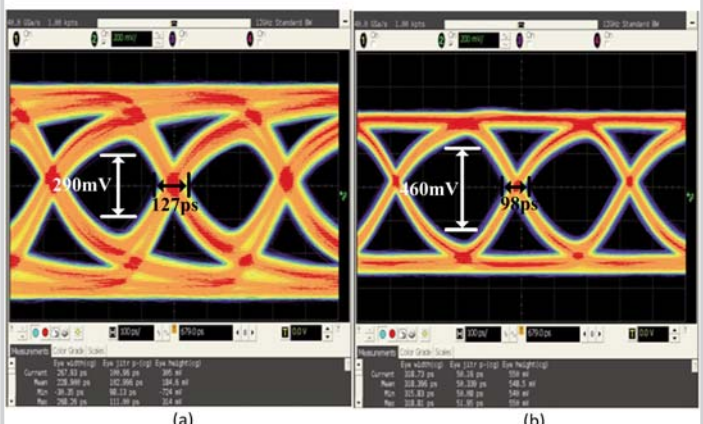

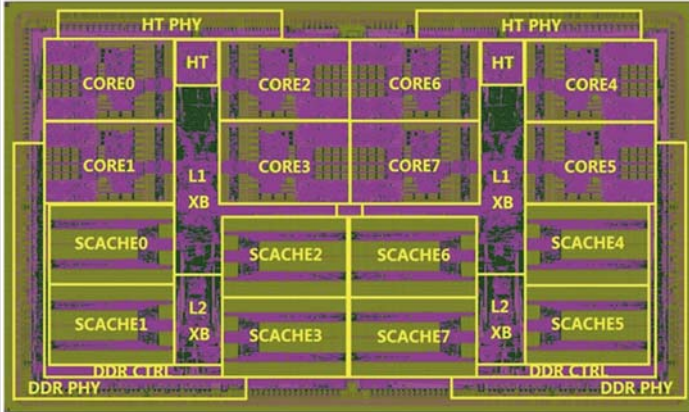**Figure 3.5.6: Eye diagram of 2.8Gb/s data over 20cm channel (a) without pre-emphasis and (b) with 3dB pre-emphasis.**

Figure 3.5.7: Die photo of Godson-3B1500.