

DYNAMIC VOLTAGE SCALING IN A FPGA-BASED SYSTEM-ON-CHIP

Jose Luis Nunez-Yanez

Vassilios Chouliaras

Jiri Gaisler

Electronic Engineering Department
Bristol University
Bristol., BS8 1UB, UK
email: eejlmy@bristol.ac.uk

Electronic Engineering Department
Loughborough University
Loughborough, LE11 3TU, UK
email: v.a.chouliaras@lboro.ac.uk

Gaisler Research
Forsta Langgatan 19
Goteborg, SE-143 27, Sweeden
email: jiri@gaisler.com

ABSTRACT

This paper presents a DVS (Dynamic Voltage Scaling) enabled SoC (System-on-Chip) processing platform based on the Leon3 open-source processor and dynamically reconfigurable clock synthesis technology available in Virtex-4 Xilinx FPGAs. A special DVS monitor unit maintains correct operation of the processor core at a given voltage by tracking the behavior of an internal delay line and stopping the processor clock through a digital clock management (DCM) macroblock when a timing error is about to occur. Upon detection of a new valid working point the DVS monitor unit reconfigures the main DCM to synthesize a new frequency-adjusted CPU clock signal and reactivates the processor. The energy savings and operation range of the technology are evaluated in the context of video coding applications by executing different motion estimation kernels.

1. INTRODUCTION

Reconfigurability, in which a single silicon platform can support multiple applications via rapid reconfiguration of the logic gates, is a topic of increasing importance. Existing and emerging consumer audio/video standards, and the need for product differentiation, result in new features that must be incorporated in already deployed devices. The mean cost of design, verification and fabrication of billion-transistor dedicated ASICs [1] is particularly high and will be increasingly difficult to justify in the face of high capacity reconfigurable platforms. It is expected that the emergence of such highly complex SoCs and FPGA technology will generate a new silicon processing paradigm in the form of a multiprocessor with dynamically reconfigurable FPGA-like processing nodes. The processing nodes could combine a small embedded processor used for the control parts of an application together with a reconfigurable fabric used to accelerate the parallel and compute-intensive aspects. Asynchronous NoCs [2] with good fault-tolerance capabilities could be used to link the processing nodes into a high performance computing array. Energy efficiency could become a major

factor limiting the success of such radical, multicore FPGA technologies since the former is expected to be a major inhibitor to arbitrary VLSI integration. For FPGA-based systems this is particularly critical since, compared with ASICs, energy consumption can be up by two orders of magnitude in the same technology [3]. At the application level there are a number of techniques available to the designer to better manage the available energy budget. Pipelining and datapath gating are two well-known techniques that help to reduce power by reducing the number of glitches or spurious transitions of logic gates associated with different path lengths of combinatorial logic [5]. Dynamic Voltage Scaling (DVS) is also a useful technique for trading latency for energy consumption in which the frequency/voltage pair is adjusted according to the application load and the power requirements of the design. Addressing the topic of power optimization in emerging multicore FPGAs, this paper uses commercially available technology to propose a possible architecture of a dynamic reconfigurable processing node and investigates how advanced FPGA features such as the Embedded Clock Management circuits can be used to support dynamic voltage scaling within the context of a video coding application.

2. DVS BACKGROUND

DVS has been applied in the field of FPGAs in [6]. In that work a series of arithmetic circuits are implemented (divider, multiplier, floating point unit, square root, etc) and an analysis conducted on the onset of I/O and logic delay errors. Differences are found depending on the variation rate of the voltage supply. Power and energy-savings data is collected at two clock frequencies, of 66 and 100 MHz, and different package temperatures.

A significant amount of research has been done using DVS in the context of embedded microprocessors for ASIC technologies. The work reported in [7] uses a sophisticated delay synthesizer circuit to accurately track the critical path delay as a function of process variation and voltage. The paper demonstrates how DVS is able to compensate for leakage due to process variations which means that some transistors are faster (lower threshold voltage, higher

leakage) and can consequently operate at a lower voltage while supporting the same clock frequency. The concept of DVS is taken one step further in Razor [8] in which the voltage is allowed to be reduced to a point in which the timing of the critical delay in the processor is not met and errors occur. Special circuitry detects the onset of these errors and recovers the pipeline with the help of delay error-tolerant flip-flops which consist of standard positive-edge trigger D-type flip-flops (which can potentially sample incorrect data) extended with negative-edge triggered shadow latches (designed to always sample correct data). That work demonstrates that removing the safety margin used in the delay lines and introducing some extra complexity by extending the flip-flops associated to the critical paths can enhance the energy savings capability of the architecture as long as the timing errors are maintained within certain limits.

3. DVS-ENABLE PROCESSING NODE

The processing node of the proposed architecture is based around the Leon3 SoC. The Leon3 is a 32-bit processor based on the SPARC (Scaleable Processor ARCHitecture) V8 architecture described in synthesizable VHDL (www.gaisler.com). The core is highly configurable and the source code is available from Gaisler Research under the GNU license. The SoC has been extended with a DVS control module attached as a new slave to the APB bus to expose its internal configuration and status registers to the programmers. Additionally, the original CLK generation module has been replaced with a new module based on the DCM_ADV primitive. The DCM_ADV is a digital clock management primitive that provides all the clock synthesis capabilities of the original DCM but it also provides access to dynamic reconfiguration features. These dynamic reconfiguration features enable the synthesis of new clock frequencies on the fly, without the need to reconfigure the remaining of the device, and are instrumental in our voltage-frequency self-tuning architecture in the process of identifying a new valid voltage-frequency working point on the basis of the core voltage available to the processor. An overview of the SoC is illustrated in Fig. 1 which also shows the ICAP core used to access the FPGA fabric directly in a related project. The section of the FPGA fabric which can be accessed through ICAP can be used to implement instruction extensions for the Leon3 processor.

4. DVS ARCHITECTURAL DETAILS

The logic dedicated to support DVS includes the CLK generation and the DVS control module. The external clock pin is supplied from an LVTTTL-type crystal oscillator that generates a 100 MHz clock signal. This signal is input into the clock generation block and the DVS control block. Two approaches to implement DVS can be identified

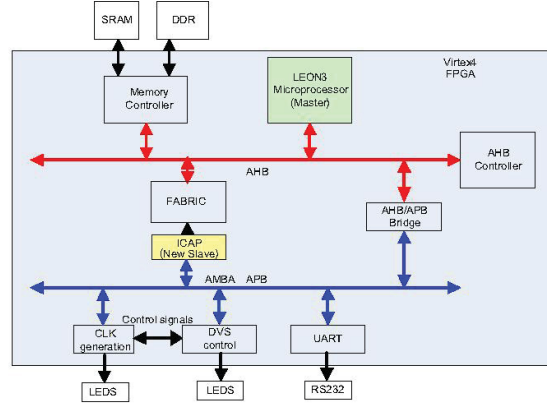
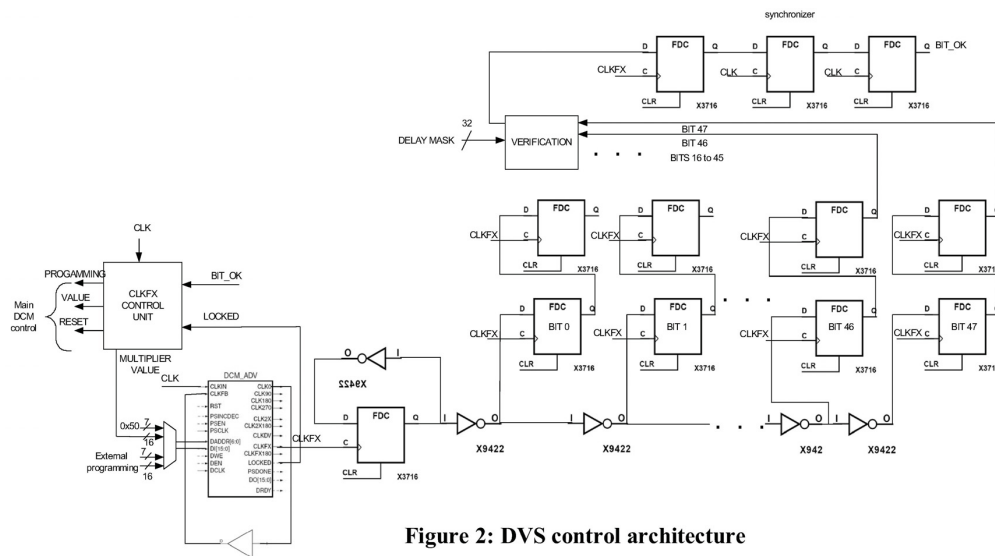


Figure 1. DVS-enable and reconfigurable Leon3 SoC

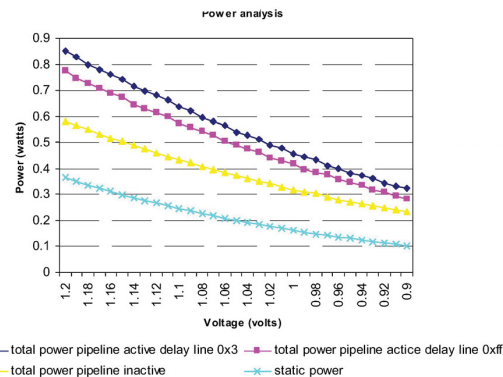
depending on which variable, voltage or frequency, changes first. The approach used in this work is designed to change the supply voltage first based on processor activity. In order to avoid the new lower voltage to cause the device to miss the critical path timing, the DVS monitor initially stops the CPU clock before starting the adaptation process. The justification is that since the critical paths are located in the processor core, extending from the integer unit to the data cache and back to the integer unit with a total of 15 levels of logic for the worst case, it is these paths which will fail first. Stopping the CPU clock prevents errors in the processor pipeline and freezes the processor state. This mechanism is designed to protect the CPU from a sudden change in voltage and/or operating conditions.

After stopping the processor core in the Leon3 SoC the control unit of the DVS monitor unit starts looking for a new valid working point, in the process reconfiguring its internal DCM_ADV primitive with new multipliers values once or several times if necessary.

We have hardwired the divider value in the DCM to the maximum permitted value of 0x1F which divides the input clock frequency by 32. Since the range of the multiplier also ranges from 1 to 32 the scheme gives a range of 32 possible different frequencies ranging from 3.15 MHz to 100 MHz (the input clock frequency from the external oscillator is 100 MHz) with a frequency step of 3.15 MHz. In this initial implementation we use the Keithley 2420 source meter to generate the variable voltage output and measure the current drawn from the VCCINT power rail. VCCINT powers the internal core and it is independent from VCCIO and VCCAUX. The wiring in the Xilinx ML402 test board has been modified and an additional jumper added so that the VCCINT pins can be connected to the original voltage regulator circuit or to the Keithley 2420 source meter unit. A simplified diagram of the DVS control module is shown in Fig. 2. The figure shows a basic delay line formed by a total of 48 inverters. The input to the inverter chain is a single bit that alternates from 0 to 1 and



propagates up to a point through the delay line before being latched by an initial chain of flip-flops. The placement of the logic cells that implement the inverters have been done using RLOC type of constraints to ensure adequate control of the delay of each stage of the inverter-based delay line. A single DCM_ADV primitive associated with the DVS control unit generates the clock signal that drives the first FF and it is also used to latch the propagated bit in the flip-flop chain. To avoid metaestability issues each flip-flop is associated with a synchronizer flip-flop placed in the same FPGA slice so that the routing delay is reduced and the possibility of the output from the second chain of flip-flops going metastable is minimized. The verification block expects to get chains of alternating 1's and 0's (or 0's and 1's depending on the cycle) to a point defined by the 32-bit Delay Mask register. The value stored in the Delay Mask register indicates how many delay line stages must be completed before a timing error is detected and a DCM reconfiguration operation started and models the delay of the critical path in the processor plus a small safety margin. The verification output is then latched into a series of FF's which are used to move from the clock domain used by the delay chain to the clock domain of the processor. The bit_OK signal is connected to an external LED and also to the CLKFX control unit. Upon detection of an error in the BIT_OK signal the CLKFX control unit will immediately reset the DCM_BASE block that generates the CLK_CPU signal that drives the processor clock.



5. ENERGY/POWER SAVINGS MEASUREMENTS

To investigate the achievable energy/power savings and correct functionality of the DVS-enhanced SoC we have selected several motion estimation kernels (hexagon, diamond and full search) extracted from an open-source

$$\text{Power} = C \cdot V^2 \cdot f + g1 \cdot V^3 + g2 \cdot V^5 \quad (1)$$

The percentage of static power to total power varies from 42% for the 1.2 volts case to 31% for the 0.9 volt case which is in agreement with equation (1) that indicates that a higher voltage has a larger impact in static power compared with dynamic power due to the different polynomial relations of power and voltage. The pipeline inactive graph shows the processor in debugging mode with instruction execution stopped and the integer unit pipeline frozen but with all the clocks running. The DVS unit has been disabled since no pipeline errors can occur in this state. The DVS unit is enabled for the other two cases using a manually-optimized delay mask at 0x3 in one case and a second case with a delay mask set at 0xff, based on data obtained from the timing analysis tool. The delay line uses 48 stages but only the upper 32 stages are compared with the delay mask. The value 0x3 indicates that as long as the control bit propagates through 19 stages no errors are detected. For the second case the control bit must propagate through 24 stages. It is apparent that the second approach will detect an error much earlier than the first and conservatively reduce the frequency of the processor to protect it from errors at a given voltage. Since the allowed clock frequency for a given voltage is higher for the 0x3 case compared with the 0xff case, it is not surprising to see that power dissipation is higher for the manually optimized delay mask. However the higher clock frequency also means that the latency of the application is lower so we need the energy analysis experiments before we can draw conclusions on efficiency. Fig. 4 shows the energy footprint of the hexagon search motion estimation kernel. The figure shows the energy needed to complete the motion estimation of a single macroblock in μ Joules. We observe that despite higher power dissipation, the manually optimized 0x3 case does a better job at finding the minimum voltage required to run the processor at particular frequency and saves energy. In this experiment the DVS control unit automatically adjusts the frequency to maintain correct operation. Overall we can see that energy consumption can be almost halved if the time budget to complete the motion estimation kernel allows it.

6. CONCLUSIONS

This paper has studied the benefits of dynamic voltage scaling implemented using the dynamically reconfigurable DCM blocks available in Virtex-4 FPGAs as an extension to a synthesizable System-on-Chip. A DVS monitor and clock generation module have been developed and the technique has been shown to successfully reconfigure the CPU clock for core voltage supply ranging from 1.2 to 0.9 volts in steps of 10 mV. The DVS block has demonstrated its ability to trade latency and energy consumption when executing several motion estimation kernels.

Macroblock motion estimation (HEX) energy savings with DVS

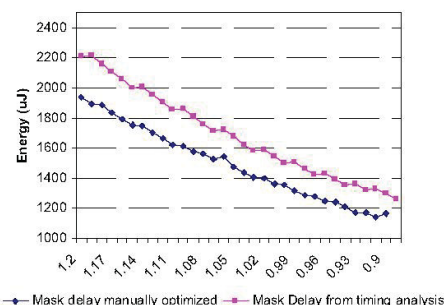


Figure 4: SoC Voltage/Energy profile

7. REFERENCES

- [1] D. Bursky, "We Must Hold the Line on Soaring ASIC Design Costs", *Electronic Design*, www.elecdesign.com, October, 2002.
- [2] W. J. Bainbridge, L. A. Plana, S. B. Furber, "The Design and Test of a Smartcard Chip Using a CHAIN Self-Timed Network-on-Chip", p. 30274, *Design, Automation and Test in Europe Conference and Exhibition Designers' Forum (DATE'04)*, 2004
- [3] P.S. Zuchowski, C.B. Reynolds, R.J. Grupp, S.G. Davis, B. Cremen, B. Troxel, "A hybrid ASIC and FPGA architecture", *IEEE/ACM International Conference on Computer Aided Design*, pp. 187-194, Nov. 2002.
- [4] Steven J. E. Wilton, Su-Shin Ang, Wayne Luk: The Impact of Pipelining on Energy per Operation in Field-Programmable Gate Arrays. *FPL 2004*: 719-728
- [5] C. T. Chow, L. S. M. Tsui, Philip Heng Wai Leong, Wayne Luk, Steven J. E. Wilton: Dynamic Voltage Scaling for Commercial FPGAs. *FPT 2005*: 173-180
- [6] S. Takahiro, et al., "Dynamic Voltage and Frequency Management for a Low-Power Embedded Microprocessor", *IEICE Trans. Electron.*, Vol. E88-C, No.4, pp. 520-527, April. 2005.
- [7] D. Shidhartha, et al., "A Self-Tuning DVS Processor using Delay-Error Detection and Correction," *IEEE Journal of Solid-State Circuits (JSSC)*, Vol 41, no.4, pp. 792-804, April 2006
- [8] Sullivan, T. Wiegand, "Video Compression - From Concepts to the H.264/AVC Standard", *Proc. the IEEE, Special Issue on Advances in Video Coding and Delivery*, Vol. 93, No. 1, pp. 18-31, January 2005.