

# High Performance State Retention with Power Gating applied to CPU subsystems – design approaches and silicon evaluation

David Flynn, Fellow, R&D  
ARM Ltd, Cambridge, UK  
david.flynn@arm.com

## ABSTRACT

Power management is of increasing concern and challenge to SOC and product designers [1], [2]. Power Gating (PG) is now well understood as a technique for reducing static leakage power when circuits are idle [3]. State-Retention Power Gating (SRPG) enhancements in hardware [4] can address fast wake-up latency and transparency to system software but have area, performance and robustness/reliability impacts that need minimizing [5].

This presentation addresses practical application of State Retention Power Gating to CPU subsystems, (but applicable to other SOC sub-systems) and covers what matters from the system and RTL designer perspective building on the EDA implementation support from UPF [6] and CPF [7] power intent.

Current EDA support for Power Gating is tuned around “logic-level” drive of power gates. The new techniques that are described and contrasted build on the multi-voltage aware tools and formats to add enhanced power gate performance as well as addressing state retention without the traditional area and timing penalties.

The work described in this paper is at an applied research phase and has been undertaken in collaboration with researchers in the Electronics and Computer Science faculty of the University of Southampton in the UK; the technology demonstrator implemented in Silicon (on a 65nm Low Leakage process) was co-developed and fabricated using the EUROPRACTICE “mini@sic” Multi-Project Wafer service with TSMC Inc as the semiconductor foundry [8].

## 1. BACKGROUND

The research group at ARM has worked for a number of years with customers and leading EDA partners to take complex 'expert' low-power industry techniques and facilitate their successful adoption for standard System-on-Chip designers and implementers. This increasingly requires the development of Physical IP components and model abstractions that support the current and evolving Multi-Voltage tools and UPF and CPF standards.

## 2. BUILDING ON BASIC POWER GATING

The multi-voltage tools support and associated power intent now prove to be a foundation for more advanced techniques to improve on the base-line power-gating and state retention support envisaged as the EDA tools were developed [9].

### 2.1 Multi-Voltage Power Gating

Industry standard “Multi-Voltage” EDA tools support logic level drive of the gate terminal of the power switches, while more expert approaches have traditionally been required to add Gate Bias to improve the off-current (ratio) of power gates [10].

A Super-Cutoff CMOS “buffered” power gate cell family with integrated level shifting has been developed to work seamlessly with standard EDA MV tool flows (shown in figure 1). Header power gates are of primary interest to facilitate simple generation of gate bias supply voltage (the core voltage rail augmented by small charge pump or regulated from a higher IO supply rail).

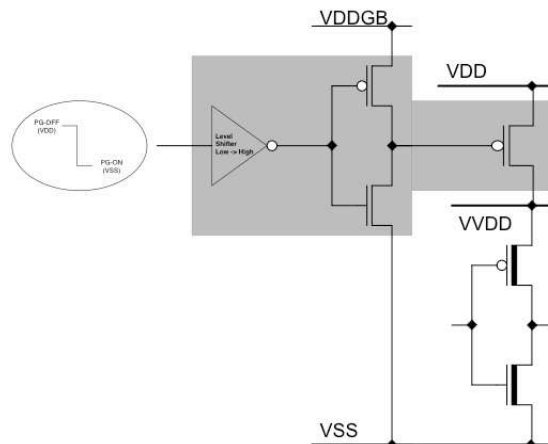


Figure 1: SCCMOS enhanced power gate

The multi-voltage internals of the enhanced switch are hidden from the implementation tools and support lower off-current with High-Vth “MTCMOS” power switches, or lower-IR drop with standard Vth switches.

## 3. ENHANCING BASIC STATE RETENTION

The experimental approach adopted has been to amortize the cost of state retention across multiple registers by splitting the power rails for high performance flip-flops (a near-zero area cost) and amortize the retention cost by managing the clamping of clocks and resets efficiently in the SOC implementation flow such that the speed and area impacts are minimized over and above the cost of Power Gating that designers well understand.

Figure 2 illustrates how the retention power domain is distributed to manage “live-slave” state retention between clock-gates and registers. For registers with asynchronous reset controls such controls must also be explicitly clamped similarly.

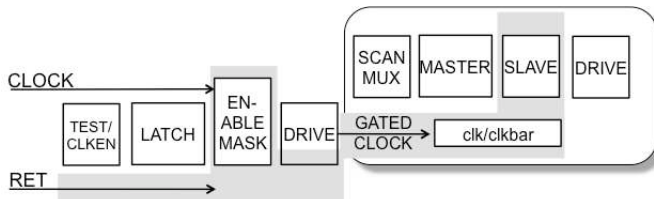
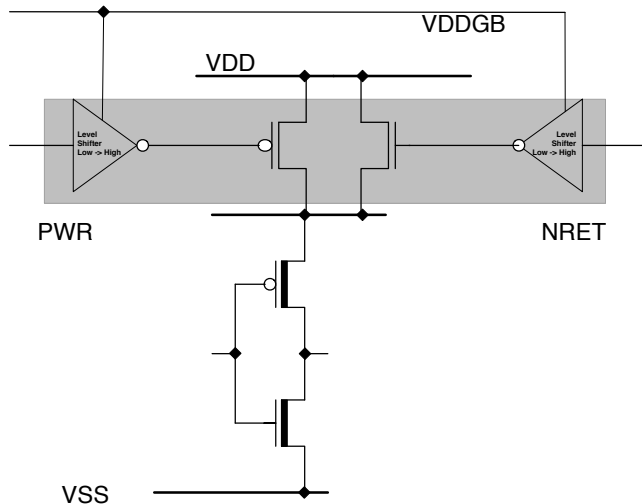


Figure 2: Advanced SRPG distributed retention domain

For short-term SRPG support the slave latches and associated clamping domains must be kept powered. For deep sleep this domain (shown with gray overlay) would be power-gated off as well (state lost PG, potentially requiring software).

Voltage scaling of the state retention rail is attractive to provided an extended SRPG mode of operation, but simple techniques such

as adding a  $V_t$ -drop that was safe at higher-voltage process nodes [4] do not provide sufficient safe state-integrity margin for latch structures on sub 90nm technologies with higher inter-device variation on latch feedback structures. Figure 3 shows the addition of a Boosted-Gate “drowsy” retention to the buffered SCCMOS power-gate of Figure 1 where the raised-voltage Gate Bias supply provides additional headroom to the scaled retention voltage.

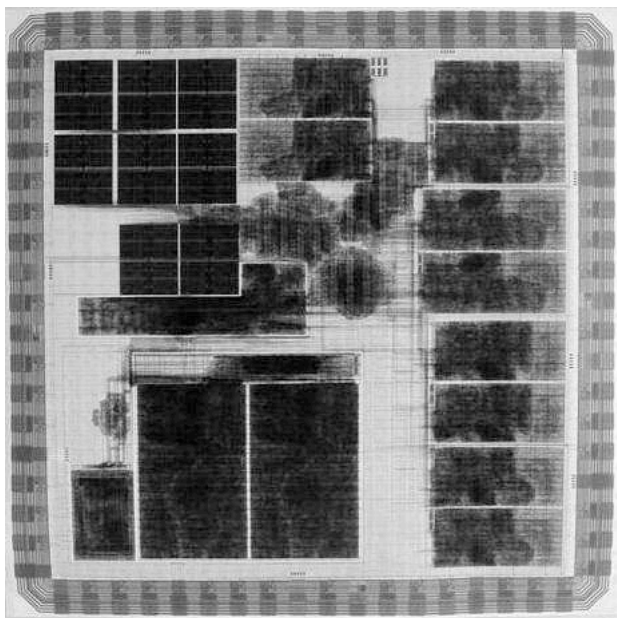


**Figure 3: SCCMOS with Boosted-Gate retention**

State retention needs to be 100% robust and reliable in the presence of power-gating transients and noise from neighboring blocks that share a common ground or supply. The underlying retention registers need to be designed to balance retention leakage power with safe retention latch structures. The poster describes the experimental structures designed and implemented to evaluate and characterize the integrity of retention registers at reduced voltages.

#### 4. TECHNOLOGY EVALUATION

Figure 4 depicts the layout of the test silicon implemented to validate the physical IP cell abstractions and EDA flow compatibility.

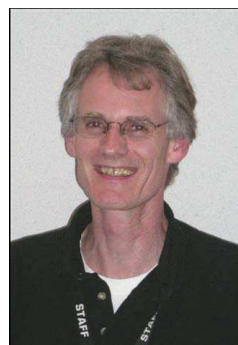


**Figure 4: Advanced ARPG test silicon (TSMC65LP)**

Due to small silicon die-size availability for rapid prototyping (2 x 2mm!) small ARM® Cortex-M0™ CPU macro-cells were chosen and constrained for performance to a worst case corner signoff at 330MHz, on the 65nm Low-Leakage technology. Five matched pairs of CPUs were instantiated with the 4 pairs on the right of the layout to evaluate standard PG and SRPG implementations plus the enhanced retention ARPG and SCCMOS plus DRPG gate bias implementations. The “tracking-pair” approach allows the implementations to be evaluated at 400MHz+ with each CPU of a pair having critical paths stressed in even and odd clock cycles, while the main SOC runs reliably zero-waits state at 200MHz. Finally, the chip includes state integrity structures in the lower-left layout to analyze state integrity and reliability in the presence of switching noise and power gating inrush.

#### REFERENCES

- [1] Mudge, Trevor, “Power: A First Class Architectural Design Constraint” IEEE Computer, vol. 34, no. 4, April 2001. <http://doi.ieeecomputersociety.org/10.1109/2.917539>
- [2] Keating, M., Flynn D. et al “Low Power Methodology Manual - for System-on-Chip Design”, Springer 2007 ISBN: 978-0-387-71818-7 <http://www.lpmm-book.org/>
- [3] Shi, K. Flynn, D. “Power Gating Design Tradeoffs and Considerations in Production Low-Power Designs”, DesignCon 2009 [http://www.designcon.com/infovault/paper.asp?PAPER\\_ID=474](http://www.designcon.com/infovault/paper.asp?PAPER_ID=474)
- [4] Mutoh S. et al. “A 1v multi-threshold voltage CMOS DSP with an efficient power management technique for mobile phone applications” ISSCC1996, pages 168–169, 1996.
- [5] Flynn, D., Gibbons, A. “Design for State Retention: Strategies and Case Studies” SNUG San Jose 2008, Track TA2
- [6] Accellera UPF Standard version 1.0, February 2007, now IEEE standard 1891 [http://www.accellera.org/activities/p1801\\_upf](http://www.accellera.org/activities/p1801_upf)
- [7] Si2 Common Power Format, CPF, specification <http://www.si2.org/?page=811>
- [8] EURO PRACTICE mini@sic programme: [http://www.europractice-ic.com/prototyping\\_minisic.php](http://www.europractice-ic.com/prototyping_minisic.php)
- [9] Kosonocky, S., “Practical Power Gating and DVFS”, Hot Chips 23 Tutorial, Aug 2011 [http://hotchips.org/uploads/hc23/HC23.17.1-tutorial1/Practical\\_PGandDV-Kosonocky-AMD.pdf](http://hotchips.org/uploads/hc23/HC23.17.1-tutorial1/Practical_PGandDV-Kosonocky-AMD.pdf)
- [10] Stan, M., “Low-Threshold CMOS Circuits with Low Standby Current,” in Proceedings of the International Symposium on Low-Power Electronics and Design. Monterey, CA: IEEE/ACM, 1998, pp. 97–99



Dr David Flynn, a Fellow in R&D at ARM Ltd, has been with the company since 1991, specializing in System-on-Chip IP deployment and methodology. He holds a BSc in Computer Science from Hatfield Polytechnic, UK and a Doctorate in Electronic Engineering from Loughborough University, UK. He is currently part-time Visiting Professor with the Electronics and Computer Science Department at Southampton University, UK. David is a primary author of the Low Power Methodology

Manual co-developed with Synopsys and launched in 2007.

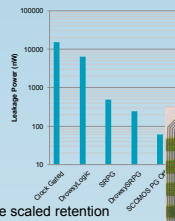
# High Performance State Retention with Power Gating applied to CPU subsystems - design approaches and silicon evaluation

David Flynn, ARM - HotChips-24, August 2012

## Building on basic Power Gating (as inferred by CPF, UPF) to address near-zero-overhead State Retention Power Gating, enhancing CPU leakage mitigation modes

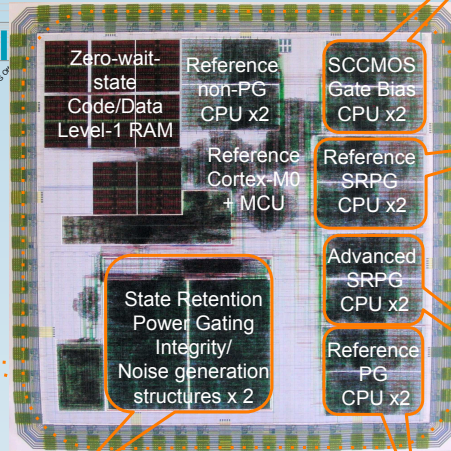
- Power Gating, PG, is well supported by Electronic Design Automation tools
    - Based on logic-level drive of Header or Footer "MT-CMOS" switches
  - State Retention Power Gating, SRPG, is also richly supported in the power intent standards
- This work builds on industry standard EDA flows to allow more expert implementation techniques to be utilized without resorting to full-custom design

- To support gate-bias (overdrive) techniques to enhance power gating
  - Super-Cutoff CMOS, SCCMOS, gate drive to power switches
  - Header power gates in this case
- To provide advanced State Retention optimized for performance and minimal area impact
  - Advanced SRPG compared to the conventional EDA preferred register level abstraction
- Combining the above to support additional leakage reduction modes including "drowsy" voltage scaled retention



## Technology demonstrator

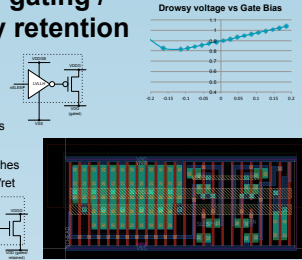
- Based on mature technology - TSMC65LP process "Tokachi-1" reference silicon
  - Including academic research (University of Southampton, UK)
  - With acknowledgement to EU EuroPractice "Mini-ASIC" research program
- Built using multiple instances of ARM Cortex8-M0 processor
  - 14 CPUs with one as a primary system MCU, 200MHz sign-off
  - Including 5 pairs optimized for performance (330MHz worst-case sign-off)
- Built using standard RTL and power intent
  - (Synopsys Inc. Multi-voltage EDA tools plus standard UPF)



## Super Cutoff power gating / Boosted-Gate drowsy retention

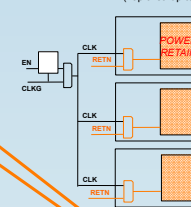
"Standard Cell" abstractions of

- Super-Cut-off Power Gating
  - Enhanced turn-off
  - Smaller Standard-Vt switches
- Boosted Gate drowsy retention
  - Very small Standard-Vt switches
  - Safer than full "diode-drop" Vret
- Clean EDA deployment



## Addressing overheads of conventional SRPG

Retention cost at every flop... (flop area up to 1.3x+) ... or shared



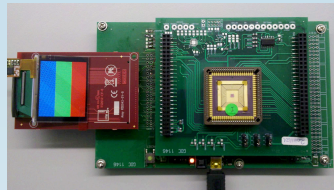
## Low-impact SRPG

- Alternative to the EDA preference for Retention abstracted per-flop
  - With the associated area and performance cost
- "Clamp" (low) the clock from the final clock-gate
  - And insert a dummy clock gate where there is no gating
  - Share the cost of clamp across local cluster of flops
  - And "live with this" more expert "flow" in order to reap the benefits
- Reset also requires clamping
  - For asynchronous-reset flops
  - (without breaking test tools...)

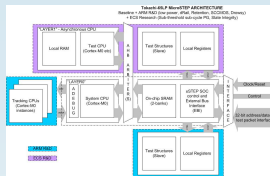
## IP Deployment Reference designs

- Building on UPF and CPF PG inference
  - Well supported in Multi-Voltage tools
- Methodology and design flow "proving"
- Representative of real-world designs
  - In terms of multi-voltage challenges
  - But quick to design/validate/fabricate/analyze
- e.g. only 5M transistors, tiny 3.5mm<sup>2</sup> area
  - But 13 power domains, 3 VDD rails
  - Analog pads to observe "virtual" rails
  - Experimental structures to analyze integrity
- Artisan® libraries plus R&D prototype cells

## Evaluation and characterization platform

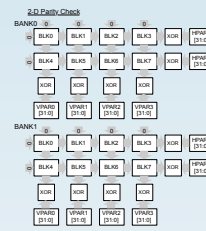


## SoC architecture



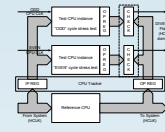
## Retention integrity analysis structures

- Reconfigurable register (synthesized) arrays
  - Controlled noise / retention
- 2-D parity analysis
  - Bank0 or Bank 1
  - With noisy other bank
- Level-shift scan chains
  - To analyze VRET scaling
- First fail detect
  - Real-time compare
  - Level shifted detect
  - for voltage sensitivity
  - X-Y of first failing bit
  - Interrupt on error (raise voltage/wake/check)



## Controlled overclocking harness

- MCU reference system uses pairs of experimental CPU macros
- Output comparators flag divergent behaviour
- Catches timing errors
- Can count cycles to error or generate interrupts



- Experimental macros clocked at double speed - 400MHz
- Fast-slow every other cycle to stay in-step with reference CPU
- Pairs required to ensure every instruction is run at-speed

## HPSRPG Physical-IP abstraction

- "Live-Slave" flops
  - Dual-rail
  - Guarded clock
- Asynchronous Reset
  - As above
  - Guarded reset
- Clock Gate cells
  - Protect the clock

