## 19.6 Voltage-Scalable Frequency-Independent Quasi-Resonant Clocking Implementation of a 0.7-to-1.2V DVFS System

Fahim U. Rahman, Visvesh S. Sathe

University of Washington, Seattle, WA

Clock power remains a substantial contributor to power dissipation, from ultra-low-power to high-performance systems [1, 2, 3]. Recently, resonant clocking has been shown to achieve power reduction in clock distribution networks [2, 3]. However, the limited voltage-frequency (VF) scalability of resonant clocking implementations remains a key drawback. Continued aggressive use of DVFS will require *efficient* resonant clocking across the increasingly wider operating range of the system. In this paper, we present a test chip implementing quasi-resonant clocking (QRC), incorporating a voltage-scalable, frequency-independent and DVFS-enabled resonant clock architecture. The test chip achieves a 32-to-47% clock power reduction over a 0.7-to-1.2V supply-voltage range.

Figure 19.6.1 illustrates existing resonant-clocking techniques and the QRC approach. Standard resonant-clocked systems employ an inductor to form an LC tank, and drive it close to its natural frequency $\omega_0$. However, away from $\omega_0$, power savings rapidly diminish, limiting frequency scalability. Intermittent resonant clocking (IRC, see Fig. 19.6.1) [4] achieves a frequency-scalable "blip" waveform. However, IRC is designed for sub-threshold designs, and reliability concerns due to non-CMOS clock voltage levels limit $V_{dd}$ scalability. Furthermore, IRC produces a fixed-width blip waveform, preventing duty-cycle control and its usage in phase paths.

The proposed QRC concept [5] relies on runtime control to alternate between resonant *Transition* and conventional *Hold* modes within a clock cycle to yield a voltage-scalable resonant clocked system with frequency-independent efficiency and complete duty-cycle control. In QRC, the resonant mode efficiently and maximally transitions the clock before switching to a conventional mode to complete the voltage transition, maintaining the clock level for an arbitrary duration thereafter. Voltage *transitions* are governed by the system natural frequency ($\omega_0$), independent of the actual operating frequency. Moving from the resonant to conventional mode involves detaching the inductor from the clock network, which must occur when no current flows through the inductor for both reliability and efficiency reasons – a major QRC architectural objective is to enable such precise timing. Voltage-scaling limitations are avoided by ensuring within-supply rail QRC clock voltages.

Figure 19.6.2 shows the QRC architecture, and relative timing of key signals. As in [2, 3], inductor L forms an LC tank with the clock network capacitance ($C_{clk}$). Coupling capacitor $C_0$ (>> $C_{clk}$) prevents DC inductor current flow. A footer switch $M_f$ controls the connection between L and $C_{clk}$ for every clock transition, allowing the system to alternate between *Hold* and *Transition* modes. At steady-state, given $C_0$ >> $C_{clk}$, $C_0$ effectively acts as a $V_{dd}/2$ voltage source (for a 50% clock duty-cycle). At the start of a clock cycle, $V_{clkPLL} = V_{clk} = 0$. As $clk_{PLL} \rightarrow V_{dd}$, f is asserted, resulting in an inductor current build-up through $M_n$ (*Build-up*). After $T_{BLD}$, $n \rightarrow 0$, turning off $M_n$, allowing an LC-driven transition of *clk* to $V_{dd}$ (*Transition*). After $T_{RE}$, (the rise edge time $\approx \pi/\omega_0$), *f* is de-asserted, disconnecting the inductor precisely when the inductor current, $I_L$=0 (or equivalently, when *clk* has maximized its resonant voltage transition). Simultaneously, $M_p$ turns on, pulling *clk* to $V_{dd}$. With L disconnected at this time, $M_p$ can hold *clk* at $V_{dd}$ indefinitely (*Hold*) until the next $clk_{PLL} \rightarrow 0$ transition. The sequence of events for *clk*→0 is similar although *p* is now de-asserted after $T_{BLD}$ and *n* after $T_{FE}$. Use of a footer device $M_f$ instead of a pass-gate [2, 3, 5] for inductor control provides improved gate overdrive, lowering resistance and power. However, because $V(C_0) \approx V_{dd}/2$, if clk=0 and *f* is de-asserted, $V_{sense} = -V_{dd}/2$ causing efficiency-degrading current backflow from *GND* to *sense*. Ensuring $M_f$ remains in cutoff requires that *f* be driven to at least $-V_{dd}/2$. The required signal timing and voltage levels for QRC signals are controlled by the timing control module (TCM) and footer driver (FD) modules.

Figure 19.6.3 shows the TCM and FD modules, which drive the *n, p* and *f* waveforms shown in Fig. 19.6.2. The TCM ensures that the events $I_L$=0 and f→0 occur simultaneously at every clock transition, thereby moving to the *Hold* mode when $I_L$=0, precisely when the clock has completed its resonant transition. Failure to align these two events raises reliability concerns due to transient ringing on *sense,* and degrades efficiency. The TCM performs a DLL operation, arbitrating between these two events with a phase-detector (a p-type Strong-ARM latch that senses the source-drain terminals of $M_f$), and adjusting $T_{FE}$ and $T_{RE}$ to appropriately delay *f*. Once lock is achieved, the phase-detector output dithers, indicative of $I_L$=0 at *f→0*. The DLL maintains lock in the event of a dynamic voltage-scaling event. Once locked, the DLL transitions to a low-bandwidth mode, reducing the TCM power to insignificant levels. MUXes controlled by $clk_{PLL}$ enable context-dependent *n, p,* and *f* signal delays in a glitch-free manner. The FD module drives *f* to $V_{dd}$ through $M_0$, $M_1$. $M_f$ is turned off by connecting its gate and source terminal, which alternates between *sense and GND* depending on the state of *clk*. When *clk=0*, $V_{sense} = -V_{dd}/2$ (Fig. 19.6.2) requiring that *f* connect to *sense* through $M_5$. When *clk=1*, $V_{sense} = V_{dd}/2$ and *f* connects to *GND* through $M_2$, $M_3$. The feedback connection of *sense* to $M_4$ prevents reverse current flow through $M_2$ and $M_3$ when $V(f) = -V_{dd}/2$. $M_1$ and $M_7$ provide gate-oxide stress protection to $M_0$ and $M_6$, while thick-oxide devices are employed for $M_4$ and $M_5$. Performance degradation from thick-oxide devices is offset by a higher gate overdrive in those devices during conduction.

Figure 19.6.4 shows the QRC test-chip architecture. Fabricated in a 9-layer 65nm CMOS process, the clock load of the QRC test chip is offered by an 8-way pipelined 32b MAC array with BIST. An on-chip ring oscillator provides the clock source ($clk_{PLL}$). The QRC system drives the MAC array clock through a 5-level H-tree. Except $M_f$ and *n/p* clock drivers, the entire design (including the clock trees) was implemented with an ASIC design flow. The clock network wire capacitance accounts for approximately half of the total clock load – an overhead comparable to modern buffer-driven conventional clock distribution.

To quantify conventional-mode power dissipation for comparison, L and $C_0$ are disconnected from the system to omit QRC-related clock loading. Resonant clock waveforms exhibit more gradual slews compared to their conventional counterparts [2, 3]. Clock driver banks were therefore disabled (gated-off) for conventional clocking to match QRC slew rates, significantly lowering clock buffer power. All clock power measurements include control logic (in the case of QRC), clock driver and the entire clock network load, while ensuring that QRC and conventional clocking slews are similar. Sample waveforms of QRC and conventional clocks with comparable slew rates are shown in Fig. 19.6.4.

Figure 19.6.5 shows measured energy-per-cycle ($E_{PC}$) vs. operating frequency for QRC and conventional clocking at $V_{dd}$=0.8V. A flat $E_{PC}$ trend across the 10-to-100MHz range demonstrates how QRC decouples efficiency from operating frequency by using resonance to enable clock *transitions*. The test chip allows for user-defined $T_{FE}$ and $T_{RE}$ (see Fig. 19.6.2) parameter settings, which override the TCM determined values. Fig. 19.6.5 shows the effect of varying $T_{FE}$ (=$T_{RE}$ for this experiment) on measured clock power dissipation, highlighting the importance of timing precision for efficient operation. Lower than optimal delays fail to take full advantage of resonance, while excess delay configurations incur a steeper power penalty due to the additional $I^2R$ power wastage on top of dissipation incurred in bringing *clk* to the supply-rails.

Figure 19.6.6 illustrates the VF scalability of the QRC test chip. The design and ring-oscillator supply voltages are scaled over 0.7V-1.2V, and $E_{PC}$ measurements are taken for both QRC and conventional clocking. The measured power reduction using QRC over conventional clocking with similar clock slew varies from 32% ($V_{dd}$=0.7V) to 47% ($V_{dd}$=1.2V). Reduced efficiency at lower $V_{dd}$ is due to increased $M_f$ resistance, which dominates $I^2R$ losses. Designs seeking lower operating $V_{dd}$, require a charge-pump to provide a boosted gate-overdrive to $M_f$ to mitigate this degradation. A die-micrograph of the test chip, and comparison to previous works are shown in Fig. 19.6.7.

*Acknowledgments:*
Sanjay Pant, Carlos Tokunaga, Shidartha Das, Jabeom Koo, Anthony Smith, Alvin Loke, Michael Khbeis, Jamal Nasser.

*References:*

[1] D. Jeon et al., "A Super-Pipelined Energy Efficient Subthreshold 240 MS/s FFT Core in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 23-34, 2012.
[2] V. Sathe et al., "Resonant-clock Design for a Power-efficient, High-volume x86-64 Microprocessor," *ISSCC Dig. Tech. Papers*, pp. 68-69, 2012.
[3] P. Restle et al., "Wide-Frequency-Range Resonant Clock with On-The-Fly Mode Changing for the POWER8™ Microprocessor," *ISSCC Dig. Tech. Papers*, pp. 100-101, 2014.
[4] H. Fuketa et al., "Intermittent Resonant Clocking Enabling Power Reduction at Any Clock Frequency for 0.37v 980khz Near-Threshold Logic Circuits," *ISSCC Dig. Tech. Papers*, pp. 436-437, 2013.
[5] V. Sathe, "Quasi-Resonant Clocking: A Run-time Control Approach for True Voltage-Frequency-Scalability," *IEEE/ACM International Symp. on Low-Power Electronics and Design*, pp. 87-92, 2014.

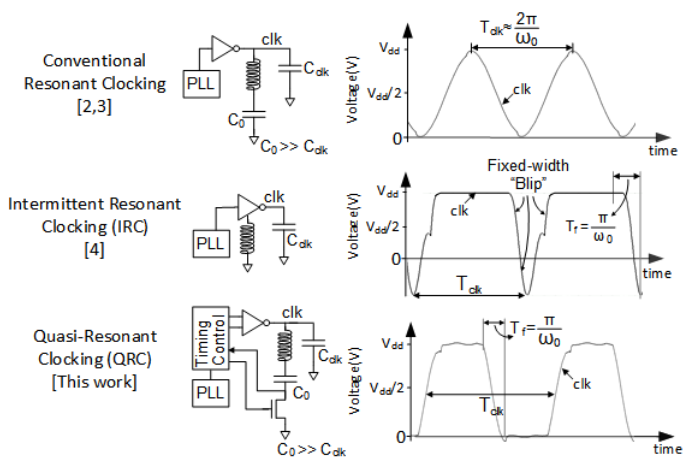



**Figure 19.6.1: Resonant clocking waveforms driving a clock load $C_{clk}$. Conventional resonant clocking is efficient close to its natural frequency. IRC provides a fixed-width "Blip" waveform that undershoots GND. QRC provides within-rail clocks and frequency-independent efficiency.**

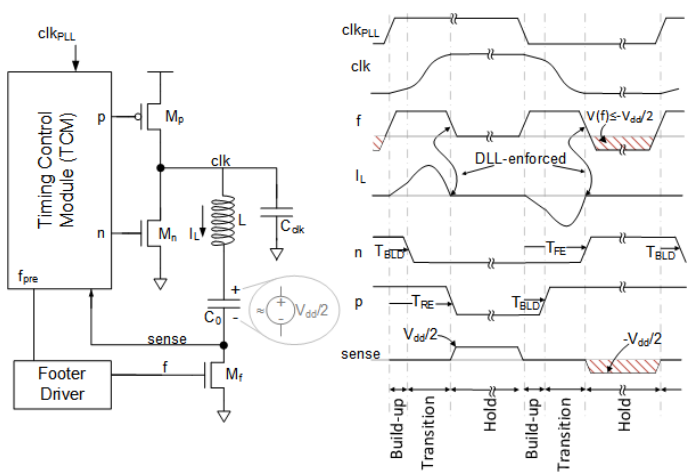



**Figure 19.6.2: QRC architecture and timing diagram of key signals. The TCM regulates the timing of p, n, and f signals in relation to the $clk_{PLL}$, and ensures that the footer is turned off precisely when $I_L$=0.**

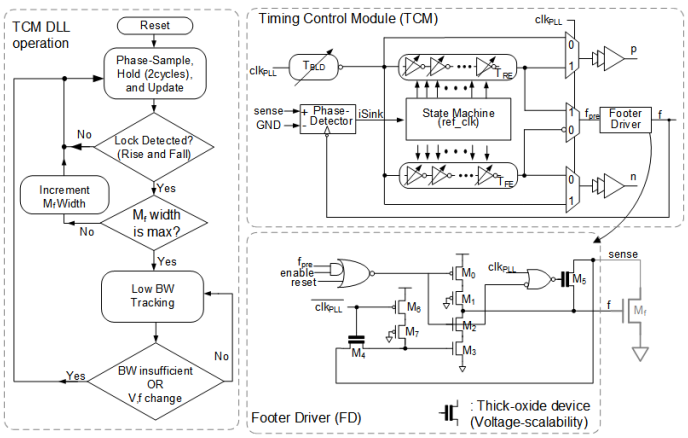



**Figure 19.6.3: TCM structure and operation. The dual-DLL aligns the de-assertion of *f* with the $I_L$=0 event (detected by sampling $M_f$ source-drain terminal polarity). A staged $M_f$ up-size mitigates reliability challenges until lock is achieved.**

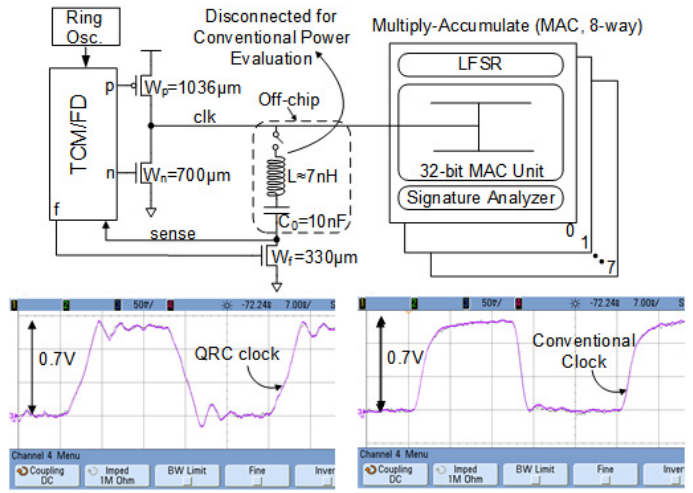



**Figure 19.6.4: QRC prototype test chip architecture, and measured conventional and QRC clock waveforms using active probes. All power comparisons are made ensuring similar conventional and QRC clock slews.**

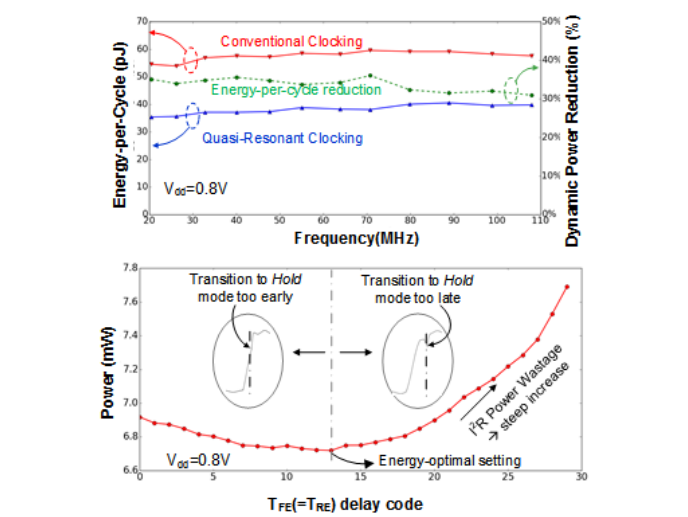



**Figure 19.6.5: Measured clock network $E_{PC}$EPC vs. operating frequency for conventional QRC, and measured power vs. TFE (=TRE) settings.**

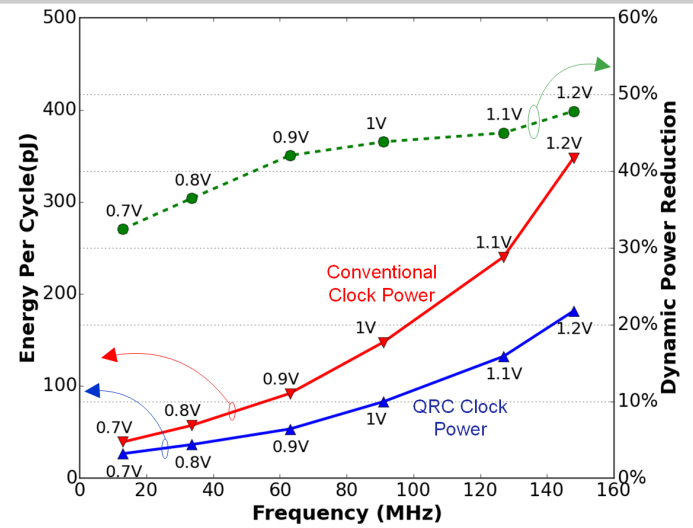
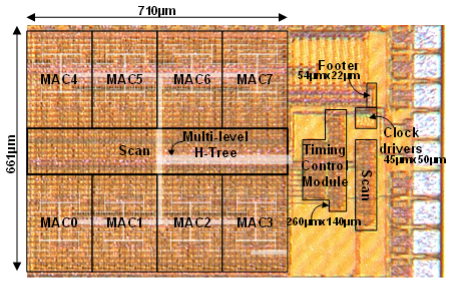



**Figure 19.6.6: Measured clock $E_{PC}$ and dynamic power reduction vs. operating frequency.**

19




| | ISSCC'12[2] | ISSCC'14[3] | ISSCC'13[4] | This work |
|---|---|---|---|---|
| Process Technology | 32nm CMOS | 22nm SOI | 40nm CMOS | 65nm CMOS |
| System Resonant frequency | 3.3GHz | ≈3.1GHz,4GHz (two modes) | Always Resonant | Always Resonant |
| Voltage-Frequency scalable resonance | No | No | No | Yes |
| Voltage Range | 1.0V-1.2V | 0.75-1.05V | 0.37V* | 0.7V-1.2V |
| Frequency Range | 2.4-4GHz | 2.5GHz - 5GHz | DC - 0.98MHz | DC – 152MHz |
| Duty Cycle Control | Limited | Limited | No | Yes |
| Dynamic Power Reduction | 15%-30% | 25%-33% | 36% | 32%-47%† |
| Inductor | On-Chip (1nH-3nH each) | On Chip (0.2nH-2.5nH each) | Off-chip 7µH | Off- chip ~7nH |

* The architecture does not allow for supply voltages greater than $V_{out}/2$ due to reliability challenges.
† Efficiency varies depending on $V_{out}$. Efficiency largely constant across frequency range (See Fig. 4)

**Figure 19.6.7: Die micrograph of the QRC prototype test chip and comparison with previous works.**


978-1-4673-9467-3/16/$31.00 ©2016 IEEE