Activity-Sensitive Flip-Flop and Latch Selection for Reduced Energy

Seongmoo Heo, Ronny Krashinsky, and Krste Asanović

***Abstract*—This paper presents new techniques to evaluate the energy and delay of flip-flop and latch designs and shows that no single existing design performs well across the wide range of operating regimes present in complex systems. We propose the use of a selection of flip-flop and latch designs, each tuned for different activation patterns and speed requirements. We illustrate our technique on a pipelined MIPS processor datapath running SPECint95 benchmarks, where we reduce total flip-flop and latch energy by over 60% without increasing cycle time.**

***Index Terms*—Clocking, flip-flops, latches, low power.**

I. INTRODUCTION

Flip-flops and latches (collectively referred to as timing elements in this paper) are heavily studied circuits, as they have a large impact on both cycle time and energy consumption in modern synchronous systems [1]–[9]. Previous work has focused on the energy-delay product of timing elements (TEs), but real designs include many TEs that are not on the critical path and this timing slack can be exploited by using slower, lower energy TEs. Instead of simultaneously optimizing for delay and energy, critical TEs should be optimized to reduce delay and noncritical TEs should be optimized to reduce energy. For example, [10] used different structures for critical and noncritical flip-flops in the context of a logic synthesis design flow.

Previous work often measured energy consumption using a limited set of data patterns with the clock switching every cycle [2]–[6], [8], [9]. But real designs have a wide variation in clock and data activity across different TE instances. For example, low-power microprocessors make extensive use of clock gating [11], [12] resulting in many TEs whose energy consumption is dominated by input data transitions rather than clock transitions. Other TEs, in contrast, have negligible data input activity but are clocked every cycle.

In this paper, we show significant energy savings when each TE instance is selected from a heterogeneous library of designs, each tuned to a different operating regime. We use detailed energy analysis to compare a number of TE designs, including designs that exploit particular combinations of signal activity and timing slack. We gather statistics on TE activity in a pipelined MIPS microprocessor running SPECint95 benchmarks and show that activity-sensitive TE selection can reduce total TE energy without increasing cycle time. To the best of our knowledge, this paper is the first work that systematically exploits *signal activity* together with timing slack to reduce TE energy by selecting different structures.

II. LATCH AND FLIP-FLOP DESIGNS

Figs. 1 and 2 present schematics for the latches and flip-flops used in this paper. We restricted our designs to fully static structures with single-rail inputs and outputs. Where TEs had complementary outputs, we loaded only the selected output. We do not penalize inverting TEs (e.g., PPCLA) because, in general, it is not obviously preferable to have either true or complement output. To ensure design robustness, we required that circuits have input buffers to isolate input sources from any actively driven feedback nodes (e.g., PTLA). Also, for each TE design we sized both low-power and high-speed versions, identified by `-lp` and `-hs` suffixes, respectively.

When choosing TEs for a real design library, other multiple factors come into play, including: input drive and output load, presence of differential inputs, desirability of complementary outputs, use of dynamic logic, robustness to clock skew and process variations, and the ability to provide time-borrowing. These factors will change the types of TE in a library, but we still expect activity-sensitive selection will help reduce energy.

Feasible TE designs are also dependent on the overall circuit layout and clocking strategies. In this paper, we target custom-designed bit-sliced datapaths in which a global clock is distributed to local clock drivers for each multibit (e.g., 32-bit) flip-flop and latch. Each local driver has a clock gating input and generates both true and inverted clock signals, so clock inverters are not required in individual TEs (except for pulse generators in some pulsed latch designs).

PPCLA [see Fig. 1(a)] is a transparent latch based on the PowerPC 603 design, which is known to be reasonably fast and energy-efficient [8]. PTLA [see Fig. 1(b)] is a pass-transistor latch, chosen for its low clock load. SSALA [see Fig. 1(c)] is a fully static differential sense amp latch, chosen for its low clock load. SSA2LA [see Fig. 1(d)] is a minor variant of SSALA, with greater clock load but lower data transition energy when clock is gated. CPNLA [see Fig. 1(e)] is PPCLA preceded by a clocked pseudo-nMOS input buffer, which reduces input data transition energy when the latch is closed. When the latch is transparent, the p-transistor in the clocked inverter acts as the pseudo-nMOS load and so dissipates considerable static power when the data input is high.

PPCFF [see Fig. 2(a)] is a master–slave flip-flop using PowerPC-style latch stages, known for low energy and delay [8]. SSAFF [see Fig. 2(b)] uses static sense-amp master–slave latch stages, chosen for low clock load. SAFF [see Fig. 2(c)] is the StrongARM flip-flop [13]. MSAFF [see Fig. 2(d)] is SAFF with a modified output stage [6] to reduce delay for higher loads.

Pulsed latch structures employ an edge-triggered pulse generator to provide a short transparency window. Compared to master–slave flip-flops, pulsed latches have the advantages of requiring only one latch stage per clock cycle and of allowing time-borrowing across cycle boundaries. The major disadvantages of pulsed latch structures are the increased susceptibility to timing hazards and the energy dissipation of the local clock pulse generators. Pulse generators can be shared among a few latch cells to reduce energy, if care is taken that the pulse shape does not degrade due to wire delay, signal coupling and noise. We measured designs both with individual pulse generators and with pulse generators shared among four latch bits, in which case we divide the pulse generator energy among the four latch instances.

HLFF [see Fig. 2(e)] operates as a pulsed transparent latch and is regarded as one of the fastest known flip-flop designs [1]. HLSFF [see Fig. 2(f)] is HLFF with a shared inverter chain. SSAPL [see Fig. 2(g)] is a pulsed version of SSALA with individual pulse generators, while SSASPL [see Fig. 2(h)] has a shared pulse generator. Note that the two series transistors in SSAPL are replaced by a single transistor in SSASPL.

Manuscript received June 20, 2001; revised October 17, 2001. This work was supported in part by Defense Advanced Research Projects Agency (DARPA) PAC/C under award F30602-00-2-0562.

The authors are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 21390 USA (e-mail: ronny@mit.edu).

Digital Object Identifier 10.1109/TVLSI.2007.902211

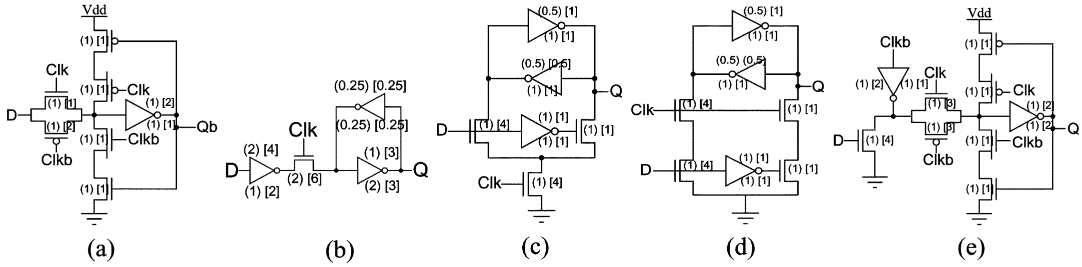

Fig. 1. High-enabled latch designs. Transistor sizes are shown for a low-power design (in parentheses: (n)) and a high-speed design (in brackets: $[n]$). A transistor labeled with size n means that its W/L ratio is n times that of a minimum-sized transistor. For gates, the sizes of all transistors are shown. (a) PPCLA. (b) PTLA. (c) SSALA. (d) SSA2LA. (e) CPNLA.

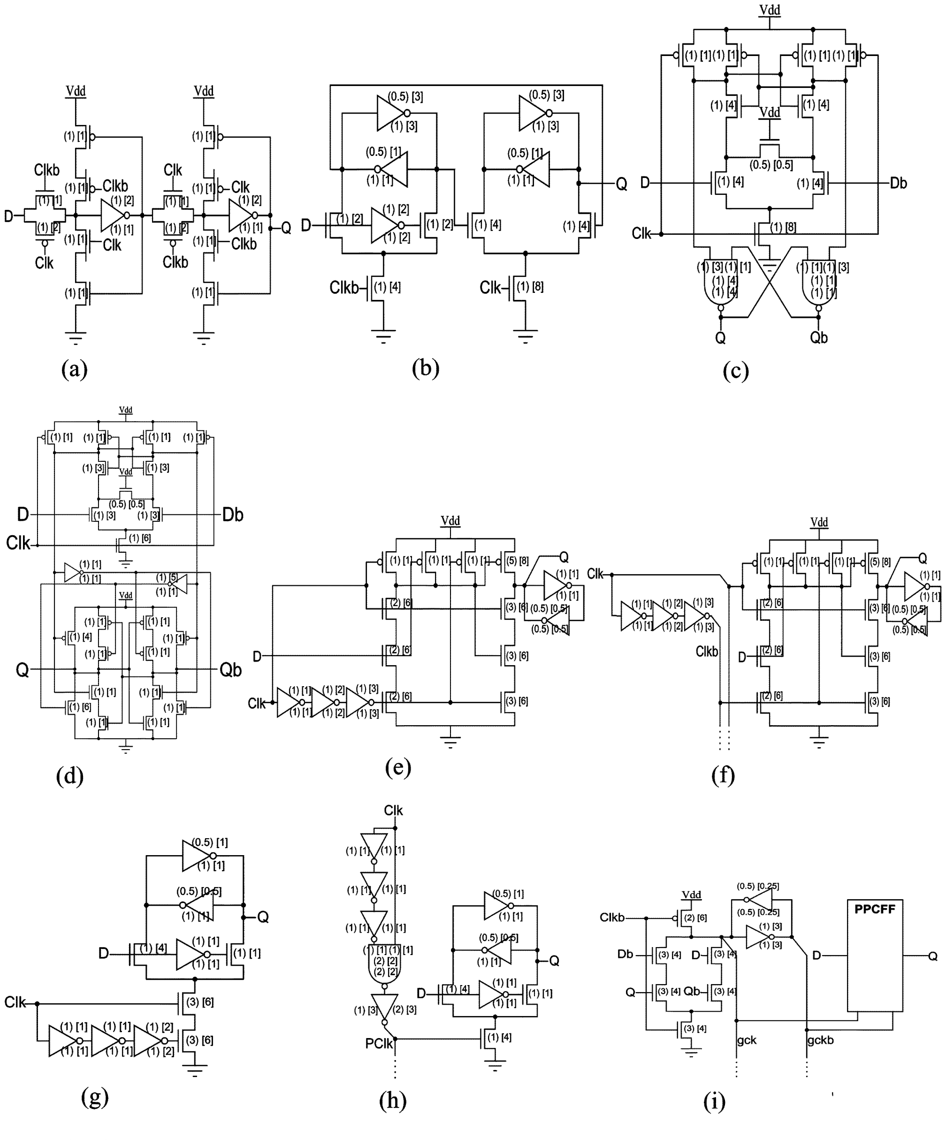

Fig. 2. Positive-edge-triggered flip-flop designs. Transistor sizes are labeled as in Fig. 1. (a) PPCFF. (b) SSAFF. (c) SAFF. (d) MSAFF. (e) HLFF. (f) HLSFF. (g) SSAPL. (h) SSASPL. (i) CCPPCFF.

Finally, CCPPCFF [see Fig. 2(i)] is a conditional clocking flip-flop based on the design presented in [9], which in turn is an improvement on [5] and [7]. The goal of this design is to reduce energy when the input data does not change by gating the clock within the flip-flop.

III. DELAY AND ENERGY CHARACTERIZATION

Our test-bench setup is similar to [8]. The data input was driven with a minimum-sized inverter which was itself driven by a loaded minimum-sized inverter to generate realistic input signals. The clock inputs were designed to simulate a local clock buffer, and the clock drivers were sized to give equal clock rise and fall times for each TE design. The TE outputs were loaded with a 7.2 fF capacitance, simulating a fanout of four minimum-sized inverters (FO4-min). Other studies [4], [6], [8] use strong input drivers and much larger output loads (200 fF). However, we extracted capacitance values for a processor datapath (described in the following) including transistor gates and drains and wire

TABLE I
DELAY FOR FLIP-FLOPS AND LATCHES

	Delay (ps)			Delay (ps)	
Flip-Flops	HS	LP	**Latches**	HS	LP
PPCFF	395	448	PPCLA	151	175
SSAFF	452	740	PTLA	252	571
SAFF	310	442	SSALA	221	424
MSAFF	288	440	SSA2LA	263	465
HLFF	204	415	CPNLA	212	260
HLSFF	204	278			
SSAPL	225	467			
SSASPL	214	487			
CCPPCFF	899	1022			

TABLE II
TE ENERGY CONSUMPTION FOR TESTS OF FIG. 3

Test:	1	2	3	4	5	6	7	8
Low-Power Flip-Flop (fJ/cycle)								
PPCFF-lp	95	97	59	**13**	202	200	145	106
SSAFF-lp	**43**	**43**	110	45	246	230	**133**	131
SAFF-lp	120	130	21	23	**196**	**194**	154	**81**
MSAFF-lp	191	190	21	23	268	267	223	117
HLFF-lp	210	361	**15**	14	380	381	329	120
HLSFF-lp	127	303	21	14	299	306	253	84
SSAPL-lp	163	165	56	68	325	310	228	138
SSASPL-lp	88	88	39	39	206	206	137	83
CCPPCFF-lp	57	57	189	59	733	691	378	218
High-Speed Flip-Flop (fJ/cycle)								
PPCFF-hs	105	106	75	14	234	233	166	127
SSAFF-hs	108	108	198	74	504	475	287	252
SAFF-hs	270	290	35	42	399	401	329	170
MSAFF-hs	383	305	31	36	461	458	394	222
HLFF-hs	370	634	29	22	591	598	541	213
HLSFF-hs	274	559	31	23	523	531	464	168
SSAPL-hs	230	233	72	102	454	418	317	187
SSASPL-hs	128	128	70	70	322	322	205	135
CCPPCFF-hs	82	105	228	57	809	765	433	269
Low-Power Latch (fJ/cycle)								
PPCLA-lp	47	46	13	**61**	**108**	**106**	77	**36**
PTLA-lp	**18**	29	32	179	203	192	113	41
SSALA-lp	22	**22**	39	101	123	139	**72**	50
SSA2LA-lp	26	25	28	109	135	132	80	41
CPNLA-lp	91	969	**9**	601	1131	631	831	55
High-Speed Latch (fJ/cycle)								
PPCLA-hs	49	49	14	57	106	103	77	39
PTLA-hs	25	54	61	172	212	204	126	73
SSALA-hs	47	47	70	141	188	242	118	94
SSA2LA-hs	33	45	40	162	201	196	120	57
CPNLA-hs	144	1734	17	1069	2008	1102	1473	89

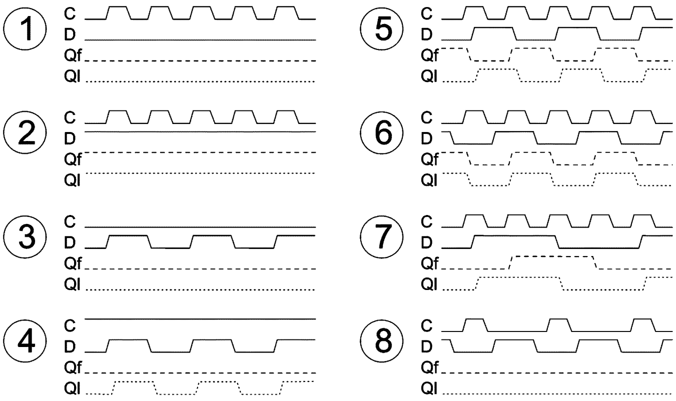

Fig. 3. Waveforms for flip-flop and latch tests. The data output waveforms are shown for a positive-edge-triggered flip-flop (Qf, dashed), and a high-enabled latch (Ql, dotted).

substrate and coupling capacitances and found that over 40% of TEs have output loads less than the FO4-min load, over 60% have loads less than twice this amount, and none with loads over 60 fF. For brevity, we consider only one size of output load, but, in general, TE characterization should consider a variety of loads [14].

TE designs were implemented in a TSMC 0.25-μm CMOS technology. Layouts were extracted using the SPACE 2-D extractor [15]. Tests were run under nominal conditions of $V_{\rm dd} = 2.5$ V and $T = 25\,^{\circ}$C. Table I shows timing for both high-speed (hs) and low-power (lp) TEs obtained using HSpice. For latches, delay is defined as the D-Q propagation delay. For flip-flops, we used the minimum D-Q delay as proposed in [8].

Traditionally, the power consumption of flip-flop and latch designs has been measured using an ungated clock and a small number of input activation patterns [2]–[6], [8], [9]. Instead, we adopt a more accurate methodology in which all possible states (e.g., clock value, input value, output value) of the TE are enumerated and the energy consumption of each state transition is measured [16]. We measured the energy consumption of each transition using HSpice and present a summary of this data in Section IV. Detailed results are available separately [17].

IV. ENERGY ANALYSIS

We constructed several example waveforms, shown in Fig. 3, to exemplify the different operating regimes for TEs. Tests 1 and 2 emphasize clock activity. Tests 3 and 4 emphasize data activity. Tests 5–7 exhibit high clock, input data, and output data activity. Test 8 has both clock and input data activity, but no output activity.

The calculated energy consumption for both high-speed and low-power TEs for these example waveforms is shown in Table II (the minimum energy for each test is shown in bold). The optimal TE for each regime varies considerably. Some designs perform extremely well in certain regimes, but extremely poorly in others. For example, in test 2 the low power SSAFF design uses eight times *less* energy than the HLFF structure, but in test 3 it uses seven times *more* energy. Another good example of a TE specialized for an operating regime is CPNLA. This latch design is by far the best choice for test 3, but by far the worst choice in all other cases.

These results also highlight the flaw in many prior TE analyses which test only a limited set of data activations with clock always ungated [2]–[6], [8], [9]. These studies typically look only at tests 5–7. The optimal TE choice may be very different, however, if tests 1–4 enter into consideration. Also, these studies have typically optimized TEs for energy-delay product. Our results show that if we size a design for high-speed and low-power separately, the energy usage can differ substantially. When the TE is not on a critical path, the low-power design should be used, and when timing is critical, the high-speed design should be used. If TEs are only optimized for energy-delay product, the result will be a slower circuit that burns more power.

V. PROCESSOR EVALUATION

To evaluate the effectiveness of designing with diverse flip-flop and latch structures, we tested our idea on a processor datapath. The design is a classic 32-bit MIPS RISC five-stage pipeline, including caches and system coprocessor registers. Aggressive clock gating is used to avoid clock transitions for the gated flip-flops and latches, and also to avoid spurious toggling of downstream functional units. The datapath contains 22 multibit flip-flops and latches, totaling 675 individual bits.

A fast cycle-accurate simulator [18] was used to count the relevant TE state transitions. The simulator tracks the input and output values of all blocks in the designs (flip-flops, adders, MUXes, etc.) and is cycle-accurate for both the high and low regions of the clock period. However, it does not accurately track the timing of signals and hence

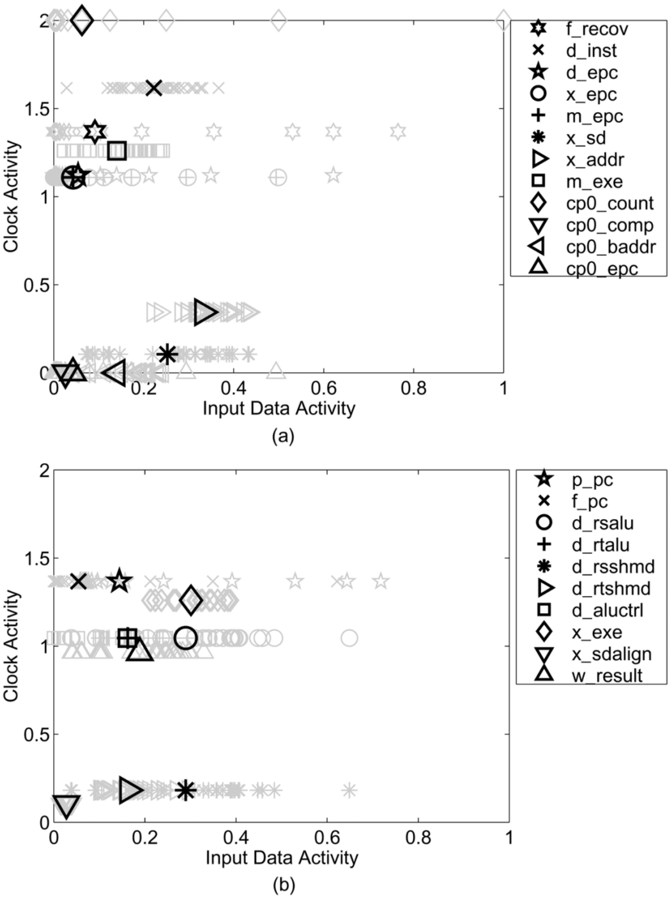

Fig. 4. Clock and input data activity (number of transitions per clock cycle) for multibit (e.g., 32-bit) registers in the CPU datapath. The black markers represent the average for each multibit flip-flop and latch, while the gray markers show the distribution of the individual bits. (a) Flip-flops. (b) Latches.

TABLE III
BREAKDOWN OF THE TOTAL TE ENERGY IN THE PROCESSOR

	Flip-flops (mJ)			
	HLFF-hs	Lowest-Energy		SSASPL-hs
f_recovpc	25.1	SSAFF-lp	**3.57**	8.12
d_inst	**31.2**	SSAFF-lp	6.52	**12.52**
d_epc	20.5	SSAFF-lp	**2.74**	6.53
x_epc	20.3	SSAFF-lp	**2.62**	6.41
m_epc	20.2	SSAFF-lp	**2.55**	6.30
x_sd	2.6	SAFF-lp	**1.06**	2.19
x_addr	**8.0**	SAFF-lp	2.57	**4.18**
m_exe	24.6	SSAFF-lp	**4.76**	9.30
cp0_count	42.6	SSAFF-lp	**4.80**	12.07
cp0_comp	0.1	HLFF-lp	**0.03**	0.16
cp0_baddr	0.3	HLFF-lp	**0.18**	0.78
cp0_epc	0.1	HLFF-lp	**0.05**	0.23
Total	195.4		31.44	68.78
Sizing	129.3			51.62
HSLE	61.5			39.05
	Latches (mJ)			
	PPCLA-hs	Lowest-Energy		
p_pc	**3.22**	SSALA-lp	2.25	
f_pc	2.95	SSALA-lp	**1.72**	
d_rsalu	**3.27**	SSALA-lp	3.16	
d_rtalu	**2.81**	SSALA-lp	2.28	
d_rsshmd	0.75	PPCLA-lp	**0.70**	
d_rtshmd	0.65	PPCLA-lp	**0.63**	
d_aluctrl	1.26	SSALA-lp	**0.97**	
x_exe	3.88	SSALA-lp	**3.65**	
x_sdalign	**0.30**	SSA2LA-lp	0.27	
w_result	**2.74**	SSALA-lp	2.42	
Total	21.84		18.06	
Sizing	21.31			
HSLE	20.02			
	TE total (mJ)			
Total	217.2		49.5	90.62
Sizing	150.6			72.93
HSLE	81.5			59.07

does not model glitches. Glitching activity would have the effect of increasing the input data activity for TEs and could possibly affect the optimal design choice. In low-power datapath designs, however, glitching activity is usually kept to a minimum.

For benchmarks, we chose five programs from SPECint95: perl(test, primes), ijpeg(test), m88ksim(test), go(20,9), and lzw (an optimized version of compress). In total, the benchmarks executed 1.71 billion instructions in 2.69 billion cycles.

Fig. 4 shows a summary of the TE state transition counts obtained from simulation, presented as overall clock and input data activity. We see that various TEs have substantially different activation patterns, and that data activity tends to be very low, while clock activation is generally much greater.

Table III shows the total TE energy breakdown in the processor datapath for the entire benchmark test set. For reference, the energy for the total datapath other than TEs was about 210 mJ for these tests. For each multibit TE, we show the energy for the fastest TE (HLFF-hs, PPCLA-hs), along with that for the lowest energy TE. We also include SSASPL-hs as a high-speed flip-flop option since it is only slightly slower than HLFF-hs (214 ps versus 204 ps) but uses much less energy. The figures in bold represent the TEs chosen when we use a high-speed-lowest-energy (HSLE) algorithm, in which a fast design is used for any timing-critical TE, and the design which results in lowest energy is used otherwise. When applying HSLE, if using slower TEs would cause a noncritical timing path to become critical, then we would use the fastest TE instead, but this did not arise in our processor design. In this study, we chose a single design for each multibit TE, and found that choosing the optimal design for each individual TE only improved results by less than 1%, as clock activity for all individual TEs in a multibit TE is identical and data activity tends to be similar.

The totals given show the energy for a fast design with homogeneous TEs, the saving achieved by transistor sizing using a homogeneous structure, and the saving using HSLE activity-sensitive selection. For flip-flops, HSLE selection reduces energy by 69% compared to a fast homogeneous design using HLFF-hs, and 52% compared to a design with transistor sizing. If we start with SSASPL-hs as the base case, the saving is 43% compared to a homogeneous design and 25% compared to a design with transistor sizing. For latches, the opportunity to save energy is reduced because they are simpler structures, and the fastest latch (PPCLA) is also quite energy efficient for the activation patterns in the datapath. Nevertheless, the energy saving with HSLE selection is 8.3% compared to a homogeneous design using PPCLA-hs, and 6.1% compared to a design using transistor sizing.

Overall, the savings we get for flip-flops and latches using HSLE activity-sensitive selection is 63% compared to a homogeneous design with HLFF-hs and PPCLA-hs and 46% compared to a design with transistor sizing. If SSASPL-hs is used as the base case flip-flop, the HSLE saving is 35% compared to a homogeneous design and 19% compared to a design with transistor sizing. Table III shows that several different TE structures are used in the optimized design, validating our hypothesis that a heterogeneous mix of TE structures can result in a lower energy design without degrading performance.

Designing with a heterogeneous mix of flip-flop and latch structures may have the disadvantage of complicating timing verification. However, advanced designs with clock gating already perform verification for each local clock independently [19] and, in this case, the added complexity is minimal. Additionally, many of the alternative TE structures are used on noncritical timing paths for which verification is usually simpler. A heterogeneous mix of TEs may also affect the glitching activity in a circuit. However, in datapaths this effect will be small since each multibit TE uses only one design, and critical TEs (for example, the ALU inputs) always use the fastest TEs available. In more irregular circuits, selecting different TEs could either increase or decrease the total glitching activity.

VI. CONCLUSION

Selecting flip-flop and latch instances from a large library of heterogeneous structures tuned for different local clock and signal activities enables a large energy saving compared to methodologies that enforce a uniform timing element structure. For a MIPS RISC processor design running SPECint95 codes, we determine that activity-sensitive selection of TEs results in a total TE energy reduction of 63% with no loss in performance compared to a high-performance design with homogeneous flip-flop and latch structures. Compared to a design which uses transistor sizing alone to reduce energy, activity-sensitive selection results in a further total TE energy reduction of 46%.

ACKNOWLEDGMENT

The authors would like to thank the numerous helpful reviewers.

REFERENCES

[1] H. Partovi *et al.*, "Flow-through latch and edge-triggered flip-flop hybrid elements," in *Dig. ISSCC*, 1996, pp. 138–139.

[2] H. Kawaguchi and T. Sakurai, "A reduced clock-swing flip-flop (RCSFF) for 63% power reduction," *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 807–811, May 1998.

[3] U. Ko and P. Balsara, "High performance, energy-efficient D flip-flop circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 8, no. 1, pp. 94–98, Jan. 2000.

[4] B. Kong, S. Kim, and Y. Jun, "Conditional-capture flip-flop technique for statistical power reduction," in *Dig. ISSCC*, 2000, pp. 290–290.

[5] T. Lang, E. Musoli, and J. Cortadella, "Individual flip-flops with gated clocks for low power datapaths," *IEEE Trans. Circuits Systems II, Analog Digit. Signal Process.*, vol. 44, no. 6, pp. 507–516, Jun. 1997.

[6] B. Nikolić, V. Oklobdžija, V. Stojanović, W. Jia, J. Chiu, and M. Leung, "Improved sense-amplifier-based flip-flop: Design and measurements," *IEEE J. Solid-State Circuits*, vol. 35, no. 6, pp. 876–884, Jun. 2000.

[7] M. Nogawa and Y. Ohtomo, "A data-transition look-ahead DFF circuit for statistical reduction in power consumption," *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 702–706, May 1998.

[8] V. Stojanović and V. Oklobdžija, "Comparative analysis of master–slave latches and flip-flops for high-performance and low-power systems," *IEEE J. Solid-State Circuits*, vol. 34, no. 4, pp. 536–548, Apr. 1999.

[9] A. G. M. Strollo, E. Napoli, and D. D. Caro, "New clock-gating techniques for low-power flip-flops," in *Proc. ISLPED*, 2000, pp. 114–119.

[10] M. Hamada *et al.*, "Flip-flop selection technique for power-delay trade-off," in *Dig. ISSCC*, 1999, pp. 270–271.

[11] V. Tiwari *et al.*, "Reducing power in high-performance microprocessors," in *Proc. DAC*, 1998, pp. 732–737.

[12] R. Gonzalez and M. Horowitz, "Energy dissipation in general purpose microprocessors," *IEEE J. Solid-State Circuits*, vol. 31, no. 9, pp. 1277–1284, Sep. 1996.

[13] J. Montanaro *et al.*, "A 160-MHz, 32-b, 0.5-W CMOS RISC microprocessor," *IEEE J. Solid-State Circuits*, vol. 31, no. 11, pp. 1703–1714, Nov. 1996.

[14] S. Heo and K. Asanović, "Load-sensitive flip-flop characterization," in *Proc. IEEE Workshop VLSI*, 2001, pp. 87–92.

[15] N. P. van der Meijs and A. J. van Genderen, "Space tutorial," Delft Univ. Technol., Delft, Netherlands, Tech. Rep. ET-NT 92.22, 1992.

[16] V. Zyuban and P. Kogge, "Application of STD to latch-power estimation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 7, no. 1, pp. 111–115, Jan. 1999.

[17] S. Heo, R. Krashinsky, and K. Asanović, "Activity-sensitive flip-flop and latch selection for reduced energy," in *Proc. 19th ARVLSI*, 2001, pp. 59–74.

[18] R. Krashinsky, S. Heo, M. Zhang, and K. Asanović, "SyCHOSys: Compiled energy-performance cycle simulation," in *Proc. Workshop Complexity-Effective Des., 27th Int. Symp. Comput. Arch.*, 2000, pp. 1–10.

[19] D. Bailey and B. Benschneider, "Clocking design and analysis for a 600 MHz alpha microprocessor," *IEEE J. Solid-State Circuits*, vol. 33, no. 11, pp. 1627–1633, Nov. 1998.