Yale University
Department of Computer Science

Fixed-Phase Retiming for Low-Power Design

Kumar N. Lalgudi[1] Marios C. Papaefthymiou[2]

YALEU/DCS/RR-1063
January 1995

[1]Department of Electrical Engineering
[2]Department of Electrical Engineering and Department of Computer Science

Fixed-Phase Retiming for Low Power Design

Kumar N. Lalgudi and Marios C. Papaefthymiou

Department of Electrical Engineering
Yale University
New Haven, CT 06520

Abstract

In this paper we describe fixed-phase retiming, a new optimization technique for the design of low power digital circuits. In fixed-phase retiming, we first transform any given edge-triggered circuit into a two-phase level-clocked circuit by replacing each flip-flop by two level-sensitive latches. Subsequently, while keeping the latches clocked on one of the phases fixed, we attempt to reduce power dissipation by relocating the remaining latches on interconnections with high glitching activity and capacitive load. Since in standard cell design the capacitance of a latch is typically smaller than the input capacitance of a combinational gate, this transformation reduces power dissipation during the opaque phase of the latch. We give a boolean quadratic programming formulation of fixed-phase retiming and describe an $O(V^4 \log V)$-time algorithm for computing a fixed-phase retiming that minimizes power dissipation, where V is the number of combinational blocks in the circuit.

1 Introduction

With the growing number of portable electronic applications, power dissipation in VLSI circuits has become a major concern. This concern has resulted in an increasing interest in the development of tools and techniques for the design of low power digital circuits. In this paper we describe an optimization technique called *fixed-phase retiming* that can help in the design of low power digital circuits. In fixed-phase retiming, a given edge-triggered circuit is transformed into a two-phase level-clocked circuit by replacing every edge-triggered flip-flop in the original circuit by two back-to-back level-clocked latches. While keeping the latches clocked by one of the phases fixed (hence the name fixed-phase), latches clocked on the other phase are relocated with a view to reduce power dissipation. Specifically, latches are placed on interconnections with high glitching activity, thereby shielding the glitches from large capacitive loads which are the primary source of power dissipation in CMOS designs. The fixed-phase retiming methodology is illustrated in Figure 1. In this paper, we present an efficient algorithm to optimize synchronous circuits for reduced power dissipation using fixed-phase retiming.

Fixed-phase retiming presents several advantages. First, since the latches clocked on one phase are kept fixed, the values of the state variables of the synchronous circuit can be obtained at the same interconnections as before optimization. Therefore, the testability characteristics of the original edge-triggered circuit remain virtually unchanged. Second, since we allow only one latch per weighted edge to move around, the final circuit is only marginally different from the original circuit. As a result, the layout doesn't change much with fixed-phase retiming. Finally, as we show in this paper, power optimization using fixed-phase retiming can be accomplished without sacrificing performance; circuit performance may, in fact, improve.

Optimization by fixed-phase retiming involves minimizing power dissipation of a given circuit by latches relocation while maintaining the clock period. This is best illustrated by

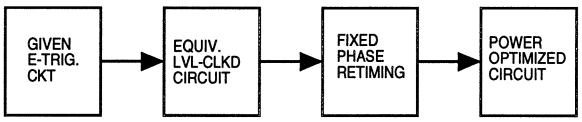

Figure 1: The fixed-phase retiming methodology for low power design.

the example given in Figure 2. Figure 2(a) shows a section of an edge triggered circuit. The numbers on the edges indicate the potential reduction in power dissipation in μW due to the presence of an edge-triggered flip-flop on that edge assuming that the rest of the circuit remains unchanged. Negative values of power reduction indicate an *increase* in power dissipation when a flip-flop is placed on an interconnection. This reduction in power dissipation can come about if the edge has a high *glitching-capacitance* product [3]. Each edge-triggered flip-flop is then replaced by two back-to-back level-clocked latches and the resulting circuit is fixed-phase retimed to result in the circuit in Figure 2 (b). If we assume a non-overlapping two-phase clocking scheme $\pi = 10 = \langle \phi_0 = 4, \gamma_0 = 1, \phi_1 = 4, \gamma_1 = 1 \rangle$ for the level-clocked latches, it can be shown that there is a reduction in power dissipation of 11.8 units. Specifically, the glitching on edges $B \stackrel{12}{\rightarrow} D$, $E \stackrel{13}{\rightarrow} F$ and $E \stackrel{-2}{\rightarrow} H$ is "masked" for 60% of the clock cycle which leads to a reduction of $0.6 \times (12 + 13 - 2) = 13.8$ units of power, while the glitching on edges $G \stackrel{10}{\rightarrow} J$ and $H \stackrel{-5}{\rightarrow} K$ is "exposed" for 40% of the clock cycle which leads to an increase in $0.4 \times (10 - 5) = 2$ units of power. Although we have made some simplifying assumptions in this example in computing power reduction, such as uniform glitching distribution over the clock period and relocation does not change glitching drastically, the example illustrates that fixed-phase retiming has a reasonable potential for reducing power dissipation in synchronous circuits.

In this paper, we show that the problem of power optimization under fixed-phase retiming can be expressed as a Boolean Quadratic Program. We then demonstrate that this problem can be reduced to a 0-1 Integer Linear Program which can be solved efficiently. Specifically, we present a polynomial-time algorithm that computes a fixed-phase retiming which maximizes the reduction in power dissipation while maintaining the performance for a given circuit and terminates in $O(V^4 \log V)$ steps where V denotes the number of combinational blocks in the circuit.

The remainder of this paper has five sections. In Section 2 we describe our graph representation model and give an overview of retiming edge-triggered circuits for low power. In Section 3, we analyze effects of fixed-phase retiming on power dissipation. Specifically, we derive a mathematical expression for the reduction in power dissipation. In Section 4 we define the power optimization problem under fixed-phase retiming and express it as a boolean quadratic program. In Section 5, we present an $O(V^4 \log V)$-time algorithm for solving the boolean quadratic program. We conclude in Section 6 with directions for further research.

2 Preliminaries

In this section we describe the graph representation of a circuit and describe power dissipation in edge-triggered circuits. We also state our assumptions about the behavior of level-clocked circuits with regard to power dissipation.

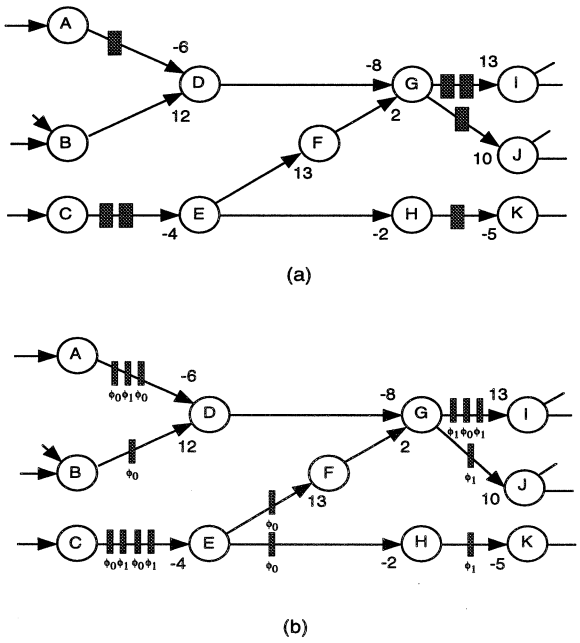

Figure 2: Section of an edge-triggered circuit that illustrates power optimization by means of fixed-phase retiming. The numbers on the edges indicate the potential reduction in power dissipation in power units when an edge-triggered flip-flop is present on that edge, assuming the rest of the network remains unchanged. (a) Initial edge-triggered circuit. (b) Fixed-Phase retimed circuit that achieves a reduction of 11.8 units of power compared to the circuit in (a) under a $\langle 4, 1, 4, 1 \rangle$ two-phase non-overlapping clocking scheme. The relocation of latches causes an increase in power dissipation on edges $E \xrightarrow{-2} H$ and $G \xrightarrow{10} J$ and a lowering in power dissipation on edges $B \xrightarrow{12} D$, $E \xrightarrow{13} F$ and $H \xrightarrow{-5} K$.

2.1 Graph representation

Given an edge-triggered circuit, we transform it into an equivalent level-clocked circuit by replacing each edge-triggered flip-flop by two level-clocked latches clocked on alternate phases by a two-phase clocking scheme $\pi = \langle \phi_0, \gamma_0, \phi_1, \gamma_1 \rangle$ as given in Figure 3. We model the resulting level-clocked circuit as a directed multigraph $G = \langle V, E, d, w, \chi, E_g, C \rangle$. The vertices V in the graph correspond to the combinational elements in the circuit. The directed edges E of the graph model the interconnections between the combinational blocks. For a combinational element v the propagation delay is given by $d(v)$ and the input phase by $\chi(v)$. If the input phase of a vertex v is $\chi(v)$, it means that $\phi_{\chi(v)}$ clocks the last latch on any path that ends at v. Each edge $u \xrightarrow{e} v \in E$ connects an output of some combinational block u to the input of another block v, and it is associated with a weight $w(e)$ that gives the latch count on the wire. Each edge $u \xrightarrow{e} v$ in the circuit graph is also associated with a pair $(E_g(e), C(e))$ where $E_g(e)$ denotes the average glitching frequency of the output of node u that leads to v and $C(e)$ denotes the capacitive load presented by node v to the output of node u. The product $E_g(e) \times C(e)$ is a measure of the power dissipation due to glitching on the edge e in the circuit.

2.2 Retiming edge-triggered circuits for low power

When an edge-triggered flip-flop is placed on a zero-weight edge, $u \xrightarrow{i} v$, there is a reduction in power dissipation since the glitching at the output of u is shielded from the rest of the circuit by the flip-flop. Assuming that the rest of the network remains unchanged, the

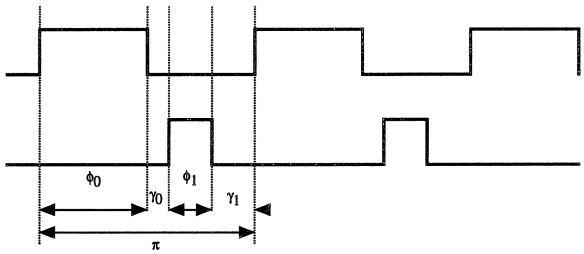

Figure 3: A two-phase clocking scheme $\pi = \langle \phi_0, \gamma_0, \phi_1, \gamma_1 \rangle$.

reduction in power dissipation is given by [3]

$$p_m^{ET}(i) = E_g(i) \times C(i) + E_g(i) \times \sum_{u \xrightarrow{j} v}^{fanout_i} (s_{j,i} \cdot C(j)) - E_g(i) \times C_{ff} \ . \tag{1}$$

Equation (1) consists of three terms. The first term $E_g(i) \times C(i)$ denotes the reduction in power dissipation at the input of v due to the masking effect. Since the glitching on edge i also propagates through its transitive fanout $fanout_i$, the masking effect of the flip flop also affects power dissipation on each edge $u \xrightarrow{j} v$ in the combinational fanout $fanout_i$. The second term $E_g(i) \times \sum_{u \xrightarrow{j} v}^{fanout_i} (s_{j,i} \cdot C(j))$ denotes the reduction in power dissipation in the edges of the transitive fanout of i, where $C(j)$ denotes the capacitive load presented by node v to the output of u. The probability that a transition on edge i propagates to edge j is denoted by $s_{j,i}$ and is given by [3]

$$s_{j,i} = Prob(j \updownarrow | i \updownarrow) \ , \tag{2}$$

where $Prob(i \updownarrow | i \updownarrow)$ denotes the probability of a transition at edge j given that there is a transition at edge i. The third term $E_g(i) \times C_{ff}$ denotes an *increase* in power dissipation due to glitching at the flip-flop inputs where C_{ff} denotes the input capacitive load of the flip-flop.

2.3 Retiming level-clocked circuits for low power

Retiming level-clocked circuits has a similar effect on power dissipation as edge-triggered retiming. Contrary to flip-flops which shield glitching for the entire clock period, level-clocked latches shield glitching for the part of the clock period that they are opaque. This does not mean, however, that fixed-phase retiming is less effective, since it can always be applied to further optimize edge-triggered circuits that have already been optimized using edge-triggered retiming sa described in [3].

Our analysis of fixed-phase retiming relies on certain simplifying assumptions. In level-clocked circuits, signals that flow through a latch during its transparent phase can initiate computations in the next combinational stage, a phenomenon termed as *cycle stealing*. As a result, data can ripple through several stages of storage elements before their propagation is complete. Our treatment of fixed-phase retiming does not take into account the effects of cycle stealing for the following reasons. First, our approach seeks to minimize the glitching component of power dissipation. Due to the inertial delay of the combinational blocks, we do not expect glitching to propagate through many combinational stages, and thus it is not an issue. Second, cycle stealing is a theoretical potential of level-clocked circuits and it is not clear how many practical circuits employ it extensively. Moreover, since our

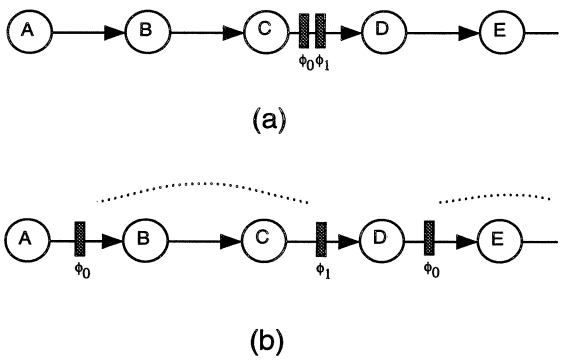

Figure 4: Masking, Exposing and Remasking. (a) A two-phase level-clocked circuit obtained by replacing edge-triggered flip-flops in the original circuit by back-to-back non-overlapping level-clocked latches. (b) Fixed-Phase retimed circuit. Moving the ϕ_0 latch from $C \rightarrow D$ to $A \rightarrow B$ masks the glitching of A from B but at the same time exposes the glitching of C to D. The incoming latch on edge $D \rightarrow E$ remasks the glitching of C from E.

methodology applies to circuits which were originally edge triggered, cycle stealing should not be significant after fixed-phase retiming. Another simplifying assumption in our analysis is that glitching is evenly distributed over the clock cycle.

3 Effects of fixed-phase retiming on power dissipation

In this section, we describe the changes in power dissipation that occur in fixed-phase retiming. We derive a mathematical expression which gives the change in power dissipation of a circuit after fixed-phase retiming.

Three things can happen to the glitching activity during fixed-phase retiming which are illustrated in Figure 4. We will assume, without loos of generality, that latches clocked on ϕ_0 are mobile and the ϕ_1 latches remain stationary. First, when a latch is placed on an edge $u \xrightarrow{e} v$ with $w(e) = 0$, the glitching of u is *masked* from v for the duration the latch is opaque. This is true for the edge $A \rightarrow B$ in Figure 4(b) where the glitching of A is not visible to B when ϕ_0 is low. Second, when a latch is removed from an edge $u \xrightarrow{e} v$ with $w(e) = 2$, the glitching at the output of u, which was previously masked from v, is now *exposed* to v for the duration the ϕ_1 latch is transparent. This is true for the edge $C \rightarrow D$ in Figure 4(b) where the glitching of C is visible to D for the duration ϕ_1 is high. Third, the exposed glitching can get *re-masked* as for the edge $D \rightarrow E$ where the glitching from C is remasked from E due to the incoming latch on edge $D \rightarrow E$.

We now present a detailed analysis of the changes in power dissipation in a circuit with fixed-phase retiming. Let E_c denote the set of edges $u \xrightarrow{e} v$ such that $w(e) = 0$, E_l denote the set of edges $u \xrightarrow{e} v$ such that $w(e) = 2$, and E_r denote the set of edges $u \xrightarrow{e} v$ such that $w(e) > 2$. We derive the power dissipation due to glitching for edges in these three sets before and after retiming. We then compute the difference of these power dissipation terms and derive a mathematical expression which denotes the reduction in glitching component of power dissipation due to fixed-phase retiming. Since fixed-phase retiming can potentially change the number of latches, the power dissipation due to the zero-delay transition activity on these latches can be different. So we compute an overall expression that denotes the reduction in power dissipation which includes the change in glitching component as well as the change in zero-delay latch power dissipation.

For an edge $u \xrightarrow{i} v \in E_c$ before retiming, power dissipation due to glitching occurs only

on the combinational blocks v and its combinational transitive fanout $fanout_i$ and is given by

$$p_{bef}(i) = E_g(i) \times \left\{ C(i) + \sum_{j}^{fanout_i} (s_{j,i} \cdot C(j)) \right\} . \tag{3}$$

The contribution to dissipation by the latches is zero since $w(i) = 0$. After retiming, there may be power dissipation by both combinational blocks and latches. The power dissipation associated with the edge $u \xrightarrow{i} v$ after retiming depends on whether the retiming process leaves a latch on i or not. When $r(v) - r(u) = 1$, i.e. a ϕ_0-latch is present on edge i, the power dissipation contributed by the latch output during the transparent phase of the latch's operation is really as though there is no latch present. This is true for $\frac{\phi_0}{\pi}$ fraction of the clock cycle. When $r(v) - r(u) = 0$, i.e. there is no latch on edge i, the power dissipation is the same as it was before retiming. As a result, the power dissipation associated with i after retiming due to dissipation in combinational blocks is given by

$$\begin{aligned} p_{aft}(i) &= \frac{\phi_0}{\pi} \cdot E_g(i) \cdot (r(v) - r(u)) \times \left\{ C(i) + \sum_{j}^{fanout_i} (s_{j,i} \cdot C(j)) \right\} \\ &+ E_g(i) \cdot [1 - (r(v) - r(u))] \times \left\{ C(i) + \sum_{j}^{fanout_i} (s_{j,i} \cdot C(j)) \right\} . \end{aligned} \tag{4}$$

The first term in Equation (4) denotes the power dissipation when $r(v) - r(u) = 1$ and the second term denotes the dissipation when $r(v) - r(u) = 0$. For the kind of level-clocked latch implementations we consider, the capacitive loads presented by a latch, whether it is open or closed, is the same. As a result, the contribution to power dissipation by the latches after fixed-phase retiming is given by

$$p^L_{aft}(i) = E_g(i) \cdot (r(v) - r(u)) \cdot C_L . \tag{5}$$

For certain other implementations of a level-clocked latch such as a pass transistor inverter combination, the input capacitance C_L may depend on whether the latch open or closed and Equation (5) maybe be differ for these. From Equations (3), (4), and (5), it follows that the reduction in power dissipation, due to the masking effect of fixed-phase retiming, associated with an edge $u \xrightarrow{i} v \in E_c$ is given by

$$\begin{aligned} \Delta p_{mask}(i) &= p_{bef}(i) - (p_{aft}(i) + p^L_{aft}(i)) \\ &= \frac{\pi - \phi_0}{\pi} \cdot E_g(i) \cdot (r(v) - r(u)) \times \left\{ C(i) + \sum_{u_j \xrightarrow{j} v_j}^{fanout_i} (s_{j,i} \cdot C(j)) \right\} \\ &+ E_g(i) \cdot (r(v) - r(u)) \cdot C_L . \end{aligned} \tag{6}$$

The power dissipation before retiming associated with an edge $u \xrightarrow{i} v \in E_l$ due to dissipation in combinational blocks is zero, since the output of the latches is a clean transition without glitches. The glitching at the input of the latches is seen by the ϕ_0-latch for the entire clock period and by the ϕ_1-latch for the duration when the ϕ_0-latch is transparent. As a result, the contribution to dissipation by the latches is given by

$$p^L_{bef}(i) = \frac{\pi + \phi_0}{\pi} \cdot E_g(i) \cdot C_L . \tag{7}$$

The power dissipation associated with edge i after fixed-phase retiming, due to dissipation in combinational blocks is given by

$$p_{aft}(i) = \frac{\phi_1}{\pi} \cdot E_g(i) \cdot r(u) \times \left\{ C(i) \cdot (1 - r(v)) + \sum_{u_j \xrightarrow{j} v_j}^{fanout_i} [s_{j,i} \cdot C(j) \cdot (1 - r(v_j))] \right\} . \quad (8)$$

When $r(u) = 0$, i.e. the ϕ_0-latch is not removed from i, then the combinational power dissipation remains unchanged and is zero. When $r(u) = 1$, Equation (8) consists of two terms. The first term is non-zero when $r(v) = 0$, i.e. no incoming ϕ_0-latch is inserted on edge i and the glitching $E_g(i)$ is visible to node v for the duration the ϕ_1 latch is open. However, when $r(v) = 1$, there is an incoming ϕ_0-latch which remasks the glitching and there is no power dissipation in the combinational fanout of i. The second term consists of the effect of glitching $E_g(i)$ on each edge $u_j \xrightarrow{j} v_j$ in the combinational transitive fanout of i which is denoted by the summation term in Equation (8). Similarly, when $r(v_j) = 0$, the glitching effect propagates, but it gets remasked when $r(v_j) = 1$. The contribution to dissipation by the latches after fixed-phase retiming is given by

$$p^L_{aft}(i) = \frac{\pi + \phi_0}{\pi} \cdot E_g(i) \cdot C_L \cdot (1 - r(u)) + E_g(i) \cdot r(u) \left\{ C_L + \frac{\phi_1}{\pi} \cdot C_L \cdot r(v) \right\} . \quad (9)$$

Equation (9) consists of three terms. When $r(u) = 0$, the power dissipation remains unchanged and is given by the first term. When $r(u) = 1$, the first term reduces to zero and the second term denotes the dissipation in the ϕ_1-latch. When $r(v) = 1$, the third term denotes the power dissipation in the incoming ϕ_0-latch on i. The term denoting power dissipation in the incoming latches on edges $u_j \xrightarrow{j} v_j$ in the combinational transitive fanout of i is not included in Equation (9) since it will be taken into account in the power dissipation term associated with edges $u_j \xrightarrow{j} v_j : w(j) = 0$. It follows from Equations (7), (8), and (9) that the change in power dissipation that arises from fixed-phase retiming for an edge $u \xrightarrow{i} v : w(i) = 2$ due to the exposing effect is given by

$$\begin{aligned} \Delta p_{expose}(i) &= p^L_{bef}(i) - (p_{aft}(i) + p^L_{aft}(i)) \\ &= \frac{\phi_1}{\pi} \cdot E_g(i) \times \left\{ (C(i) - C_L) \cdot r(u) \cdot r(v) + \sum_{u_j \xrightarrow{j} v_j}^{fanout_i} [s_{j,i} \cdot C(j) \cdot r(u) \cdot r(v_j)] \right\} \\ &\quad - E_g(i) \cdot r(u) \times \left\{ \frac{\phi_1}{\pi} \cdot [C(i) + \sum_{u_j \xrightarrow{j} v_j}^{fanout_i} (s_{j,i} \cdot C(j))] + \frac{\phi_0}{\pi} \cdot C_L \right\} . \end{aligned} \quad (10)$$

Now in addition to the reduction in the glitching component of the power dissipation, there could be a change in power dissipation due to a change in the number of latches in the circuit. This component of power dissipation is entirely due to the change in the power dissipation of clock ϕ_0 and is given by

$$\Delta p_{clk}(\phi_0) = E(\phi_0) \cdot C_{L-\phi_0} \cdot \left\{ \sum_{u \xrightarrow{e} v \in E} (r(v) - r(u) + \frac{w(e)}{2}) \; - \; |E_l| \right\} . \quad (11)$$

Equation (11) consists of two terms. The first term denotes the power dissipated in the latches in the retimed circuit and the second term represents the power dissipation in the latches in the original circuit by the ϕ_0 clock line. The switching activity on the ϕ_0 clock line is given by $E(\phi_0)$ and the capacitive load presented by the latch to the ϕ_0 clock line is given by $C_{L-\phi_0}$.

It follows from Equations (6), (10), and (11) that the net reduction in power dissipation by means of fixed-phase retiming is given by

$$\mathcal{PR} = \sum_{e \in E_c} \Delta p_{mask}(e) + \sum_{e \in E_l} \Delta p_{expose}(e) + \Delta p_{clk}(\phi_0) \ . \tag{12}$$

Thus the power optimization problem by fixed-phase retiming is equivalent to maximizing the objective function $\mathcal{PR}$ while maintaining the performance of the circuit.

4 Fixed-phase retiming via boolean quadratic programming

In this section we define the power optimization problem under fixed phase retiming. We show that the problem is a boolean quadratic program with *monotone* inequalities. It is the monotonicity of the constraints that allows us to develop an efficient algorithm for fixed-phase retiming. While optimizing circuits for power dissipation, it is desirable to maintain the performance of the circuit and ensure that the circuit still operates at the same clock speed after fixed-phase retiming. As a result, we need to impose timing constraints during the optimization process. The following lemma gives the necessary and sufficient conditions for a retiming such that a retimed circuit G_r is properly timed by a given clocking scheme π.

Lemma 1 (Lemma 35, [5]) *Let $G = \langle V, E, d, w, \chi \rangle$ be a two-phase, level-clocked circuit, let $\pi = \langle \phi_0, \gamma_0, \phi_1, \gamma_1 \rangle$ be a clocking scheme, and let $r : V \to \mathbf{Z}$ be a retiming function. Then, the retimed circuit G_r is properly timed by π if and only if for every edge $u \xrightarrow{e} v \in E$, we have*

$$r(u) - r(v) \leq w(e) \ , \tag{13}$$

and for every path $u \overset{p}{\rightsquigarrow} v$, we have

$$\begin{aligned} d(p) \ \leq \ & \pi \left(\frac{1 + w(p)}{2} \right) + \phi_{\chi(u)} && (14) \\ & + \pi \left\lfloor \frac{r(v)}{2} \right\rfloor + (r(v) \bmod 2)(\gamma_{\chi(u)} + \phi_{1-\chi(u)}) && (15) \\ & - \pi \left\lfloor \frac{r(u)}{2} \right\rfloor - (r(u) \bmod 2)(\gamma_{\chi(u)} + \phi_{\chi(u)}) \ , && (16) \end{aligned}$$

if $\chi(u) \neq \chi(v)$, and

$$\begin{aligned} d(p) \ \leq \ & \pi \left(\frac{2 + w(p)}{2} \right) + \phi_{1-\chi(u)} && (17) \\ & + \pi \left\lfloor \frac{r(v)}{2} \right\rfloor + (r(v) \bmod 2)(\gamma_{1-\chi(u)} + \phi_{\chi(u)}) && (18) \\ & - \pi \left\lfloor \frac{r(u)}{2} \right\rfloor - (r(u) \bmod 2)(\gamma_{\chi(u)} + \phi_{\chi(u)}) \ , && (19) \end{aligned}$$

if $\chi(u) = \chi(v)$.

Thus power optimization problem with timing constraints leads us to the following mathematical formulation.

Definition 2 (Power Optimization Problem - PO) *Let $G = \langle V, E, p, w, \chi, E_g, C\rangle$ be a synchronous circuit and let $\pi = \langle \phi_0, \gamma_0, \phi_1, \gamma_1\rangle$ be a two-phase clocking scheme. Let E_l denote the set of edges $u \xrightarrow{e} v \in E : w(e) = 2$ and let E_c denote the set of edges $u \xrightarrow{e} v \in E : w(e) = 0$. Moreover, let p be the least-weight path from u to v in G. The power optimization problem under fixed-phase retiming is to compute a retiming $r : V \rightarrow \{0, 1\}$ such that we*

$$\max\ (\mathcal{PR})$$

subject to the constraints

$$r(u) \leq r(v) + w(e) \qquad \text{for all } u \xrightarrow{e} v \in E\ ,$$

$$\begin{aligned} d(p) \leq\ & \pi(\tfrac{2+w(p)}{2}) - \gamma_{1-\chi(u)} \\ & + r(v)(\gamma_{1-\chi(u)} + \phi_{\chi(u)}) \\ & - r(u)(\gamma_{\chi(u)} + \phi_{\chi(u)}) \qquad \text{for every path } u \overset{p}{\rightsquigarrow} v\ . \end{aligned}$$

Expression (12) denotes the reduction in power dissipation under fixed-phase retiming which we wish to maximize. The constraints in Definition 2 are required to maintain the performance of the circuit. These are derived from Lemma 35 in [5] which gives the necessary and sufficient conditions for a retiming r, so that the retimed level-clocked circuit is properly timed by a general two-phase clocking scheme π. It is straightforward to show that for $r \in \{0, 1\}$, the conditions in Lemma 1 reduce to the constraints in Definition 2.

We now show that Problem PO has some special properties which we will exploit in Section 5 to obtain a polynomial-time algorithm for solving it. From Equation (10), it is evident that $\mathcal{PR}$ contains the quadratic terms $r(u) \cdot r(v)$ and $r(u) \cdot r(v_i)$ and thus it is a quadratic programming problem. Moreover, since $r \in \{0, 1\}$ the power optimization problem PO is an instance of boolean quadratic programming which is $\mathcal{NP}$-complete in general. We make two important observations. First, we notice that all constraints in Definition 2 are monotone inequalities (inequalities of the form $ax_i - bx_j \leq c$ where $a, b \geq 0$ are termed monotone [2]) with at most two variables per inequality. Second, if we assume that the latch capacitance is smaller than the capacitance of combinational blocks, that is, $C_L \leq C(i)$ forall $i \in E$, then all quadratic terms in the objective $\mathcal{PR}$ have positive coefficients. Specifically, the term $(C(i) - C_L) \cdot r(u) \cdot r(v)$ in Equation (10) can only take positive values with this assumption. We performed a comprehensive comparison of the input capacitances of the cells in the CMOS3 Cell Library [1] to check whether this was indeed a reasonable assumption to make. With the exception of a high impedance inverter whose input capacitance was comparable, all other combinational blocks had higher input capacitances than a level-clocked latch. Thus, the power optimization problem under fixed-phase retiming reduces to a boolean quadratic program with monotone inequalities and an objective to be maximized that has positive quadratic coefficients.

5 Polynomial-time algorithm for fixed-phase retiming

In this section we show that the boolean quadratic program of Section 4 is efficiently solvable. From a general boolean quadratic program we derive a boolean quadratic program

with the special structure that all constraints are monotone linear inequalities and the objective consists of quadratic terms with only positive coefficients. Based on the results of Hammer *et. al.* in [4], we reduce it to a boolean linear program. We then show how to solve the boolean linear program efficiently based on the results of Hochbaum and Naor in[2]. Specifically, we show that any boolean quadratic program with m constraints and n variables, whose constraints are monotone linear inequalities and whose objective consists of quadratic terms with only positive coefficients can be solved in $O(mn^2 \log n)$-time. We then give a polynomial-time algorithm that optimizes power dissipation by fixed-phase retiming and terminates in $O(V^4 \log V)$ steps, where V is the number of combinational blocks in a circuit.

The general boolean quadratic program with m constraints and n variables is defined as follows.

Definition 3 (Boolean Quadratic Programming - BQP) *Given* $\mathbf{c_1} \in \mathbb{R}^n$, $C_2 \in \mathbb{R}^{n \times n}$ *and* $\mathbf{x} \in \mathbb{B}^n$, *we wish to maximize*

$$\mathbf{c_1 x} + \mathbf{x}^T C_2 \mathbf{x} \ , \tag{20}$$

subject to

$$A\mathbf{x} \leq \mathbf{b} \ , \tag{21}$$

where $A \in \mathbb{R}^{m \times n}$ *and* $\mathbf{b} \in \mathbb{R}^m$.

We consider a special case of BQP in which the constraints are *monotone* linear inequalities and all quadratic terms in the objective have positive coefficients which is defined as follows.

Definition 4 (Boolean Monotonic Quadratic Program - BMQP) *Given* $x_i, x_j \in \mathbb{B}$, $d_{ij} \in \mathbb{R}$ *for all* $i, j = 1, \cdots, n$, *and* $P = \{(i,j) : 1 \leq i, j \leq n;\ d_{ij} > 0\}$, *we wish to maximize*

$$\sum_i^n d_i x_i + \sum_{(i,j) \in P} d_{ij}\, x_i \cdot x_j \ , \tag{22}$$

subject to

$$a_k x_i - b_k x_j \leq c_k \ , \qquad k = 1, \cdots, m \ , \tag{23}$$

where, $a_k, b_k \in \mathbb{R}$ *and* $a_k, b_k \geq 0$.

Introducing new 0-1 variables y_{ij} associated with the quadratic terms $x_i \cdot x_j$ in Definition 4 and constraining them to take the value of the product of the two literals in these terms, we arrive at a linearized form of BMQP which is defined as follows.

Definition 5 (Linearized Boolean Monotonic Quadratic Program - LBMQP) *Given* $x_i, y_{ij} \in \mathbb{B}$, $d_{ij} \in \mathbb{R}$ *for all* $i, j = 1, \cdots, n$, *and* $P = \{(i,j) : 1 \leq i, j \leq n;\ d_{ij} > 0\}$, *we wish to maximize*

$$\sum_i^n d_i x_i + \sum_{(i,j) \in P} d_{ij}\, y_{ij} \ , \tag{24}$$

subject to

$$y_{ij} - x_i \leq 0 \qquad (i,j) \in P \ , \tag{25}$$

$$y_{ij} - x_j \leq 0 \qquad (i,j) \in P \ , \tag{26}$$

$$a_k x_i - b_k x_j \leq c_k \qquad k = 1, \cdots, m \ , \tag{27}$$

where $a_k, b_k \in \mathbb{R}$ *and* $a_k, b_k \geq 0$.

POFPR(G, π)
1 Compute objective $\mathcal{PR}$ and constraints $\mathcal{C}$ of the form of PO
2 Linearize $\mathcal{PR}$ to obtain an LBMQP formulation by replacing every term of the form $r(u) \cdot r(v)$ by the variable r_{uv} such that $\mathcal{C} \leftarrow \mathcal{C} \cup \{r(u) \geq r_{uv}\} \cup \{r(v) \geq r_{uv}\}$
3 Compute r that maximizes $\mathcal{PR}$ subject to $\mathcal{C}$ using the Hochbaum-Naor technique.
4 **return** r.

Figure 5: Algorithm POFPR for solving the power optimization problem using fixed-phase retiming. The algorithm determines a retiming $r : V \rightarrow \{0, 1\}$ that minimizes the power dissipation and terminates in $O(V^4 \log V)$ steps.

It was shown in Theorem 1.7 in [4] that unconstrained boolean quadratic expressions of the form of Expression (22) and their linearized forms of the form of Expression (24) have the same optimum value. We state a corollary of that result for boolean monotonic quadratic programs in the following lemma.

Lemma 6 *The optimum solution of Problem* LBMQP *is equal to the optimum solution of Problem* BMQP.

Proof. The proof follows from a case analysis of the values taken by x_j and x_i and examining the expression $x_i \cdot x_j$ and the variable y_{ij} for each case. When $(x_i, x_j) = \{(0,0), (0,1), (1,0)\}$, $x_i \cdot x_j = 0$; moreover, Inequalities (25) and (26) ensure that $y_{ij} = 0$. When $(x_i, x_j) = (1,1)$, $x_i \cdot x_j = 1$, but y_{ij} is unconstrained. Since the coefficient of y_{ij} is positive, any solution that seeks to maximize the objective will try to force y_{ij} to be 1. For $(x_i, x_j) = (1,1)$, y_{ij} is unconstrained and consequently it will be set to 1 in the optimum solution. Thus $x_i \cdot x_j$ and y_{ij} take identical values in the optimum solution of Problems BMQP and LBMQP respectively. Thus the optimal solution of LBMQP is also the optimal solution of BMQP □

The following lemma gives the running time for computing the integer optimal solution of a monotone system of linear inequalities with bounded variables and only two variables per inequality with respect to an arbitrary linear objective.

Lemma 7 (Theorem 3.7, [2]) *The integer optimal solution of a monotone system of inequalities with respect to an arbitrary linear objective can be computed in pseudo-polynomial time, in* $O(m(\sum_{i=1}^{n} |V_i|)^2 \log(\sum_{i=1}^{n} |V_i|))$ *where* m *is the number of inequalities,* n *is the number of variables, and* V_i *is the set of integers that are contained between the largest and smallest integer feasible values of variable* x_i.

We now give the running time for solving Problem BMQP in the following theorem.

Theorem 8 *The optimal solution to the Problem* BMQP *can be computed in* $O(mn^2 \log n)$ *time where* m *is the number of constraints and* n *is the number of variables.*

Proof. From Lemma 6, it follows that the optimal solution to Problem BMQP and its linearized form LBMQP are the same. Problem LBMQP is a monotone system of linear inequalities where the variables can take only two values in the set $\{0, 1\}$ which implies that $|V_i| = 2$ for any variable x_i in LBMQP. Consequently, the running time for Problem BMQP follows immediately from Lemma 7. □

We now present an efficient algorithm for solving Problem PO which is based on the results of Hochbaum and Naor in [2]. The algorithm, given in Figure 5, first computes the objective and the set of constraints to obtain a boolean quadratic program. This boolean quadratic program is then linearized to obtain an instance of LBMQP which is then solved using the algorithm due to Hochbaum-Naor [2]. We conclude this section with the following theorem.

Theorem 9 *Algorithm* POFPR *solves Problem* PO *in* $O(V^4 \log V)$ *steps.*

Proof. We have $O(V^2)$ constraints for the fixed-phase retiming problem and $O(V)$ variables. Since Problem PO can be formulated as an LQBMP, the running time for POFPR follows immediately from Theorem 8. □

6 Conclusion

In this paper we have investigated fixed-phase retiming for designing low power digital circuits. We have shown that the ensuing optimization problem can be expressed as a boolean quadratic program which can be solved in $O(V^4 \log V)$ steps. We are currently evaluating the effectiveness of our technique. Our preliminary experiments with a 4-bit carry-lookahead adder indicate power savings of about 15%. Our experiments reveal that in addition to shielding highly capacitive nodes from glitching, fixed-phase retiming reduces power dissipation by equalizing arrival times of signals at gate inputs, thereby reducing the glitching component itself.

References

[1] D. V. Heinbuch. *CMOS3 Cell Library.* Addison Wesley, 1988.

[2] D. S. Hochbaum and J. Naor. Simple and fast algorithms for linear and integer programs with two variables per inequality. *SIAM J. Computing*, 23(6):1179 – 1192, December 1994.

[3] J. Monteiro, S. Devadas, and A. Ghosh. Retiming sequential circuits for low power. In *Digest of Technical Papers of the 1993 IEEE International Conference on CAD*, pages 398–402, November 1993.

[4] P. Hansen P. L. Hammer and B. Simeone. Roof duality, complementation and persistancy in quadratic 0-1 optimization. *Mathematical Programming*, 28:121–155, 1984.

[5] M. C. Papaefthymiou. *A Timing Analysis and Optimization System for Level-Clocked Circuitry.* PhD thesis, Massachusetts Institute of Technology, September 1993. Available as MIT/LCS/TR-605.