

Power-Driven Global Routing for Multi-Supply Voltage Domains

Tai-Hsuan Wu, Azadeh Davoodi, and Jeffrey T. Linderth*

Department of Electrical and Computer Engineering

*Department of Industrial and Systems Engineering

University of Wisconsin at Madison, USA

Email: {twu3, adavoodi, linderth}@wisc.edu

Abstract—This work presents a method for global routing (GR) to minimize interconnect power. We consider design with multi-supply voltage, where level converters are added to nets that connect driver cells to sink cells of higher supply voltage. The level converters are modeled as additional terminals during GR. Given an initial GR solution obtained with the objective of minimizing wirelength, we propose a GR method to detour nets to further save the interconnect power. When detouring routes via this procedure, overflow is not increased, and the increase in wirelength is bounded. The power saving opportunities include: 1) reducing the area capacitance of the routes by detouring from the higher metal layers to the lower ones, 2) reducing the coupling capacitance between adjacent routes by distributing the congestion, and 3) considering different power-weights for each segment of a routed net with level converters (to capture its corresponding supply voltage and activity factor). We present a mathematical formulation to capture these power saving opportunities and solve it using integer programming techniques. In our simulations, we show considerable saving in an interconnect power metric for GR, without any wirelength degradation.

I. INTRODUCTION

Power consumption is a primary design objective in many application domains. Dynamic power still remains the dominant portion of the overall power spectrum. Design with Multi-Supply Voltage (MSV) allows significant reduction in dynamic power by taking advantage of its quadratic dependence on the supply voltage.

Dynamic power is dissipated in combinational and sequential logic cells, clock network, and the (remaining) local and global interconnects. We refer to the latter as interconnect power. Interconnect power can take a significant portion of the dynamic power spectrum. For example, the contribution of the interconnect power is reported to be around 30% of dynamic power for a 45nm high performance microprocessor synthesized using a Structured Data Paths design style and about 18% of the overall power spectrum [1].

The interconnects are complex structures in nanometer technologies that span over many metal layers. The power of a route segment depends on its width, metal layer, and spacing relative to its adjacent parallel-running routes. These factors determine the area, fringe, and coupling capacitances which impact power. Furthermore, in MSV designs, the power of a routed net depends on its corresponding supply voltage. For example, a route will have lower power if all its terminal-cells have the (same) lower supply voltage. If a net connects a driver cell of lower voltage to a sink cell of higher voltage, its route includes a level converter (LC) and is decomposed into two segments of low and high supply voltages, corresponding to before and after the LC.

We propose a global routing (GR) method that optimizes the interconnect power in MSV designs. Figure 1 shows a generic design

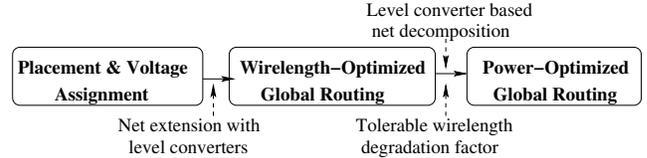


Fig. 1. Overview of GR with MSV.

flow for a MSV-based GR. After placement and voltage assignment, the location and supply voltage of each cell are known. The supply voltage is determined either through voltage island generation [2], [3], or through a row-based assignment in a standard cell methodology. Furthermore, LC(s) are added to any net that connects a driver cell to a set of sink cells of higher supply voltage. Next, GR is applied to minimize the overall wirelength (WL), where the LCs are also included as terminals of a net.

For a given WL-optimized GR solution, we propose to further detour the nets in order to optimize the interconnect power. The interconnect power can be approximated during GR since at this stage the metal layers of each route segment are known. Furthermore, the spacing of parallel routes can be estimated from the routing congestion. Given a WL-optimized solution, the nets can be rerouted to trade off WL with power. For example nets from higher metal layers can be routed to the lower ones for less wire widths and area capacitance. Nets can also be rerouted to spread the congestion, thereby increasing their spacing for less coupling capacitance. Activity factor and supply voltage can be incorporated as a power-weight for each route segment.

We present a mathematical formulation for MSV-based GR to minimize power, and present integer programming-based techniques to solve the formulation. As part of power saving, our methods spread the routing congestion and ensure no additional overflow (of routing resources) and a bounded degradation in WL compared to the initial solution.

To the best of our knowledge, this is the first work of power-driven global routing in MSV designs. Recently the work [4] discusses power-driven GR, however it does not consider the MSV case. Also, it relies on the availability of *power-efficient candidate routes* for each net but generates such candidate routes quite heuristically. As part of the contributions of this work, we show a formal procedure to generate power-efficient candidate routes from the initial WL-optimized solution while taking into account the overall WL degradation.

The remainder of the paper is divided into 4 sections. Section II describes our MSV-based interconnect model. Section III discusses our formulation and solution procedure for power minimization. Simulation results are presented in Section IV, and conclusions are offered in Section V.

This research was supported by National Science Foundation under Grant 0914981.

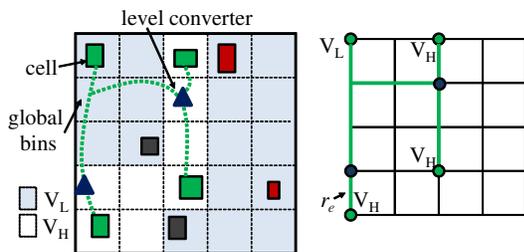


Fig. 2. Global routing model with LCs.

II. INTERCONNECT MODELING

In this section, we discuss an MSV-based GR model. We assume the level converters (LCs) are placed for some of nets and the supply voltage of each cell is known.

A. Interconnect Modeling in MSV Designs

We are given a grid-graph $G = (\mathcal{V}, \mathcal{E})$ model of the GR problem, where each vertex $v \in \mathcal{V}$ corresponds to a global bin containing a number of cells. Each edge $e \in \mathcal{E}$ represents the boundary of two adjacent bins. A capacity r_e is associated with each edge e , reflecting the maximum number of routes that can pass between two adjacent bins. A net $i \in \{1, \dots, N\}$ is identified by its terminal cells, which are a subset of the vertices \mathcal{V} . In MSV-based GR, the terminals of a net may also be the LCs. During GR, a Steiner tree t_i in G is found for each net i to connect its terminals. The length of t_i is taken to be its wirelength (WL).

Figure 2 shows an example. The chip is divided into regions. Each region has either a low (V_L) or high (V_H) supply voltage. A routed net is specified in the figure. The net has one driver terminal with V_L voltage and three sink terminals of V_H voltage. The route includes two LCs which are also considered as additional terminals of the net.

For power-driven MSV-based GR, we first decompose a net which contains a LC into a set of sub-nets. We reroute each sub-net as an individual net during power optimization. Consequently, we have $N_d > N$ number of nets after decomposition. For example, in Figure 2, the initial route is shown with its LCs. The net is decomposed into three sub-nets, each of which will be rerouted. The first sub-net connects the driver terminal in V_L to the two LCs. The second one connects one LC to one V_H terminal. The third one connects the other LC to the other two V_H terminals.

The decomposition of each net is done using its initial route and the location(s) of its LC(s). For a net containing LC, starting from its driver terminal, a sub-net corresponding to a low supply voltage is formed that connects the driver terminal to a set of LCs and/or a set of sink terminals of the same supply voltage. Next, one or more sub-nets are formed that connect the LCs to the sink terminals of the same (and higher) voltage level. Our net decomposition procedure finds a minimum number of sub-nets for each net that contains a LC such that each sub-net has only one corresponding supply voltage. Space constraints prohibit us from providing complete details of the decomposition method. The general decomposition procedure is similar to existing works [5], [6].

B. Power Modeling

Each decomposed net $i \in \{1, \dots, N_d\}$ has a corresponding supply voltage V_i and switching activity α_i . The required interconnect power for a GR solution is estimated as

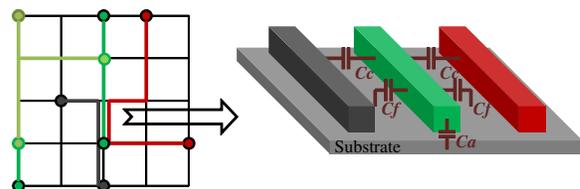


Fig. 3. Modeling route capacitance on a GR edge.

$$P = f_{clk} \times \left(\sum_{i=1}^{N_d} \alpha_i V_i^2 (C_i^{sink} + C_i^{route}) \right), \quad (1)$$

where f_{clk} is the frequency. As seen in Equation 1, the capacitance of routed net i is the sum of the capacitances of its sink cells (denoted by C_i^{sink}) and of its route (denoted by C_i^{route}). Here C_i^{sink} is a constant that does not depend on the re-routing, so it is excluded from the optimization. Note that the power of the LCs are considered fixed and thus also not considered as part of the interconnect power optimization. The capacitance C_i^{route} for a routed net i is the sum of the capacitances of its unit-length edges that are contained in route t_i (given by notation $e \ni t_i$):

$$C_i^{route} = \sum_{e \ni t_i} C_e^u. \quad (2)$$

The parameter C_e^u is the capacitance of one routed edge $e \in \mathcal{E}$. This capacitance is a function of the metal layer l_e , wire width w_e and wire spacing s_e of the edge e . Specifically,

$$C_e^u = Ca(l_e, w_e) + 2Cf(l_e, w_e, s_e) + 2Cc(l_e, w_e, s_e), \quad (3)$$

where Ca and Cf are the area and fringe capacitances with respect to substrate, and Cc is the coupling capacitance. As indicated, these capacitances are functions of wire length, width, and spacing, and are provided by the technology library through a lookup table.

In this work, we assume that only one (and a different) wire width is associated with each metal layer, so we exclude the parameter w_e , and for each edge $e \in \mathcal{E}$, its metal layer l_e is known. The spacing for edge e is estimated from the edge utilization u_e in a GR solution. Given the utilization u_e and the length of edge e , (computed from the chip dimension and the routing grid granularity), the spacing s_e is calculated to allow maximum spacing between its corresponding routes. Figure 3 shows an example for $u_e = 3$. This simple *averaging* strategy may be adjusted if more information is available at the GR stage; (e.g., the adjustment may be due to the fixed short nets which fall inside a single global routing bin). With this approximation, we can express the capacitance of a unit-length route-edge in terms of the edge's metal layer and its utilization. The total capacitance of edge e is given by the product of the per-unit capacitance C_e^u and the utilization u_e : $C_e = C_e^u \times u_e$.

Figure 4 (left) shows the curves representing area, fringe, and coupling capacitances for metal layer 1 with respect to edge utilization for a 45nm library [7], assuming each GR edge is 2μ . The summation of the 3 capacitances (C_e^u) is shown on the right.

III. POWER-DRIVEN MSV-BASED GR

In this section, we first present a mathematical formulation of power-driven MSV-based GR. We then discuss integer programming-based techniques to obtain high-quality solutions to the formulation.

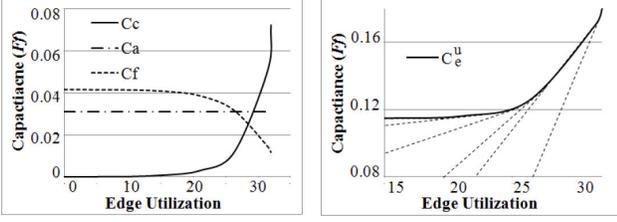


Fig. 4. Dependence of three types of capacitance on edge utilization in metal layer 1.

A. Mathematical Formulation

As described in Section II-B, the per-unit capacitance of an edge e (C_e^u) is a function of its metal layer and the edge utilization. Typically, this function is a convex increasing function, as depicted in Figure 4. We represent the function C_e^u by a set of line segments denoted by Q_e^u . For example, the set Q_e^u is composed of 7 line segments in the library used in this work [7]. Each line segment $q \in Q_e^u$ is of the form $m_q^u + r_q^u u_e$, for a given range of u_e , where m_q^u and r_q^u are derived from the library for that range. For each of the 8 metal layers in our library, the curve C_e^u is represented as 7 piecewise-linear segments.

Since the per-unit capacitance is convex, its value may be expressed in our mathematical optimization problem for GR with the following set of linear inequalities:

$$m_q^u u_e + r_q^u \leq C_e^u, \forall q \in Q_e^u. \quad (4)$$

For a given edge utilization u_e , the corresponding C_e^u is obtained from the line equation that gives the largest value of $m_q^u u_e + r_q^u$ for $q \in Q_e^u$.

To model GR we are given a routing grid graph $G = (\mathcal{V}, \mathcal{E})$, a set of decomposed multi-terminal nets denoted by N_d , and edge capacities r_e . Let \mathcal{T}_i be a collection of all Steiner trees that can route net i . We later discuss how to approximate \mathcal{T}_i by generating a set of power-efficient candidate trees with consideration of WL degradation. Each tree $t \in \mathcal{T}_i$ is associated with a binary decision variable x_{it} which is equal to 1 if and only if it is selected to route net i . Let the parameter a_{te} be equal to 1 if tree t contains edge e (if $e \ni t$). The GR problem for power minimization is given by:

$$\min_{x, s, C^u} \sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} \alpha_i V_i^2 (\sum_{e \ni t} C_e^u) x_{it} + \sum_{i=1}^{N_d} M s_i \quad (\text{PGR})$$

$$\begin{cases} \sum_{t \in \mathcal{T}_i} x_{it} + s_i = 1 & \forall i = 1, \dots, N_d \\ \sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} a_{te} x_{it} \leq r_e & \forall e \in \mathcal{E} \\ m_q^u (\sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} a_{te} x_{it}) + b_q^u \leq C_e^u & \forall e \in \mathcal{E}, \forall q \in Q_e^u \\ \sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} w_{it} x_{it} \leq W_0(1 + \beta) \\ s_i \geq 0 & \forall i = 1, \dots, N_d \\ x_{it} = \{0, 1\} & \forall i = 1, \dots, N_d, \forall t \in \mathcal{T}_i. \end{cases}$$

The first term in the expression of the objective function is the interconnect power as explained in Section II-B. It includes activity α_i and voltage V_i of net i . The capacitance of a route t of net i is obtained by adding the unit edge capacitances C_e^u for all the edges $e \ni t$. Here the route $t \in \mathcal{T}_i$ will be selected for net i only if $x_{it} = 1$.

The first set of constraints selects at most one route for each net. The slack variable s_i is equal to 1 if net i cannot be routed, and the variable is penalized in the objective function by a large parameter M

to maximize the number of routed nets. The term $\sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} a_{te} x_{it}$ represents the edge utilizations u_e . The second set of constraints ensures that the edge utilizations are within the given edge capacities. The third set of constraints determines the per-unit edge capacitance C_e^u for each edge e from its utilization, using the discussed piecewise linear model. The fourth constraint ensures the new wirelength is within a factor β of the initially-provided wirelength W_0 . Here w_{it} denotes the wirelength of route t of net i .

The constraints of formulation (PGR) are all linear. However, the objective expression is nonlinear (due to the multiplication of variables x_{it} and C_e^u). We handle the nonlinearity in a heuristic manner using a two-phase approach. First, we choose a re-routing that attempts to minimize the total capacitance of all edges. Next, per-unit capacitances are estimated (and fixed) based on the solution of the first phase, and a re-routing is sought that minimizes the total estimated power. Each of these two phases become integer *linear* programs (IPs) which are discussed in the next subsections.

B. Phase1: Minimizing Total Capacitance

Using the piecewise linear approximation for the per-unit capacitance C_e^u given by Equation (4), we may also approximate the total capacitance as

$$C_e = C_e^u \times u_e \geq m_q^u u_e^2 + r_q^u u_e \quad \forall q \in Q_e^u.$$

This (convex) nonlinear expression may be re-linearized, resulting in another piecewise linear expression for the total edge capacitance that may be used in our linear integer program for minimizing the total capacitance.

$$C_e \geq m_q u_e + r_q \quad \forall q \in Q_e. \quad (5)$$

1) *Formulation:* The formulation of phase 1 is given by the following IP:

$$\min_{x, s, C} \sum_{e \in \mathcal{E}} C_e + \sum_{i=1}^{N_d} M s_i \quad (\text{PGR-P1})$$

$$\begin{cases} \sum_{t \in \mathcal{T}_i} x_{it} + s_i = 1 & \forall i = 1, \dots, N_d \\ \sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} a_{te} x_{it} \leq r_e & \forall e \in \mathcal{E} \\ m_q (\sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} a_{te} x_{it}) + b_q \leq C_e & \forall e \in \mathcal{E}, \forall q \in Q_e \\ \sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} w_{it} x_{it} \leq W_0(1 + \beta) \\ x_{it} = \{0, 1\} & \forall i = 1, \dots, N_d, \forall t \in \mathcal{T}_i \\ s_i \geq 0 & \forall i = 1, \dots, N_d. \end{cases}$$

The objective expression is similar to formulation (PGR) but the first term is replaced by $\sum_{e \in \mathcal{E}} C_e$ which represents an estimate of the total interconnect capacitance. The third set of constraints is also updated; the variable C_e replaces C_e^u in the previous formulation, and the coefficients in the piecewise linear model are updated to use Equation 5.

2) *A Price-and-Branch Solution Procedure:* We approximately solve the (PGR-P1) using the a two-step heuristics. First, a *pricing* procedure is used to generate a set of candidate routes for each net that are power-efficient while considering the WL degradation. The pricing step approximates \mathcal{T}_i in the formulation to contain a small set of power-efficient candidate routes, instead of all the potential routes of net i . Second, *branch-and-bound* is applied to solve (PGR-P1), selecting one route for each net from the set of generated candidate routes. The standard branch and bound algorithm can be carried out using a commercial solver. This two-step procedure of generating candidate routes and then running branch and bound is commonly

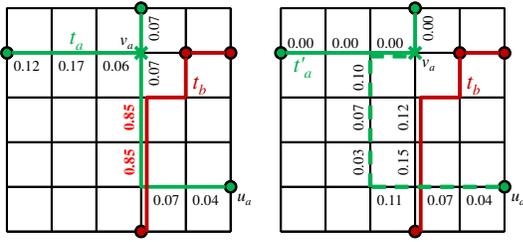


Fig. 5. Power-aware route generation.

known as price-and-branch [8], [9]. The price-and-branch procedure was recently applied to solve the GR problem for WL improvement [10]. We apply the same procedure for power improvement. The major technical difference in our procedure is in the pricing step to find power-efficient candidate routes, which we next discuss in detail.

3) *Overview of Pricing for Route Generation:* We solve a linear-programming relaxation of (PGR-P1) by replacing the binary requirements on the variables x_{it} with constraints $0 \leq x_{it} \leq 1 \forall i, \forall t$. The linear program is solved by an iterative procedure known as column-generation [11]. In column generation, we start by replacing \mathcal{T}_i (set of all possible routes of net i) in formulation (PGR-P1) by subset $\mathcal{S}_i \subset \mathcal{T}_i$, initially containing one candidate route per net. We then gradually expand \mathcal{S}_i , adding new routes that may decrease the objective function. Adding the new candidate routes is via a *power-aware pricing condition* for each net.

Before explaining the procedure in more detail, we first give the following notations:

1. We refer to the LP relaxation of (PGR-P1) in which \mathcal{T}_i is replaced by \mathcal{S}_i and $0 \leq x_{it} \leq 1$ by the “restricted master problem” denoted by (RMLP-P1); the solution of (RMLP-P1) for a given \mathcal{S}_i is denoted by $(\hat{x}, \hat{s}, \hat{C})$;
2. We refer to the dual of the restricted master problem by (D-RMLP-P1). The solution of (D-RMLP-P1) consists of $(\hat{\lambda} \leq M, \hat{\pi} \leq 0, \hat{\mu} \geq 0, \hat{\theta} \leq 0)$, corresponding to the dual variables for the first, second, and third set of constraints in the relaxed (PGR-P1), respectively.

The iterative column generation procedure including the pricing condition is enumerated below:

1. For each net $i = \{1, \dots, N_d\}$, initialize \mathcal{S}_i with one route. (In this work we start with the solution of [5]).
2. Solve (RMLP-P1), yielding a primal solution $(\hat{x}, \hat{s}, \hat{C})$ and dual values $(\hat{\lambda}, \hat{\pi}, \hat{\mu}, \hat{\theta})$ in (D-RMLP-P1).
3. Generate a new route t^* for net $i = \{1, \dots, N_d\}$. Using the solution of step 2, evaluate the pricing condition: If $\hat{\lambda}_i > \sum_{e \in t^*} \sum_{q \in Q_e} m_q \hat{\mu}_{eq} - \sum_{e \in \mathcal{E}} (\hat{\pi}_e + \hat{\theta})$, then $\mathcal{S}_i = \mathcal{S}_i \cup \{t^*\}$.
4. If an improving route for some net i was found in step 3, return to step 1. Otherwise, stop—the solution $(\hat{x}, \hat{s}, \hat{C})$ is an optimal solution to (RMLP-P1).

Step 3 gives the pricing condition in terms of the solution of the dual problem (D-RMLP-P1) obtained at the current iteration. This step can determine for a given new route t^* , if it should be added to the set \mathcal{S}_i to reduce the objective of (RMLP-P1). However, it does not specify how a new route should be found such that the pricing condition gets satisfied. We discuss a convenient graph-based procedure to generate new route t^* which satisfy the pricing condition.

4) *Route Generation for One Net:* To find improving routes for net i , we associate a weight w_e for edge e in the GR grid as:

$$w_e = \max_{q \in Q_e} (m_q \hat{\mu}_{eq}) - \hat{\pi}_e - \hat{\theta}. \quad (6)$$

By the theory of linear programming, for each edge e , at most one dual variable $\mu_{eq}, q \in Q_e$ will be positive in an optimal solution to (D-RMLP-P1). Thus, considering route t^* , we can compute the pricing condition as $\hat{\lambda}_i > \sum_{e \in \mathcal{E}} w_e$. We take advantage of this interpretation to identify promising route t^* which satisfies the pricing condition. Given a route $t \in \mathcal{S}_i$ obtained from previous iterations, we obtain t^* by rerouting branches of t with the updated edge weights so that the overall weights of rerouted branches are reduced.

We explain the procedure with the example of Figure 5. Considering two nets a and b , suppose we are initially given the routes t_a and t_b for these two nets. After step 2 at the first iteration of column generation, we obtain edge weights which are given in the figure on the left. To obtain a new route t'_a for net a , we reroute different branches of t_a . For each terminal, we identify a branch as the segment connecting it to the first Steiner point on t_a . We then reroute this branch by solving Dijkstra’s single-source shortest path algorithm [12] on the weighted graph with the weights of the first iteration, similar to [13], [14]. The route t'_a is shown in the right figure. After adding t'_a to \mathcal{S}_a we proceed to the second iteration and obtain new edge weights which are shown in the right figure.

The discussed pricing procedure is similar to [10]. However, it differs in the pricing condition and the way edge weights are set up. For solving (RMLP-P1) and its dual at each iteration we use the solver CPLEX 12.0. After obtaining the final set \mathcal{S}_i , again we use CPLEX 12.0 for the branch and bound step to get the final solution. We further accelerate the process by applying a simple problem decomposition that we will discuss in Section III-D.

C. Phase2: Considering Activity and Voltage

At phase 2, we approximate the per-unit edge capacitances using the solution from phase 1, and re-route the nets to minimize an approximation of the total power. Since the utilization (and hence capacitance) corresponding to the routing solution of phase 2 may be different from phase 1, we heavily penalize any mismatch in our optimization.

1) *Formulation:* We compute the following quantities after phase 1:

1. We define a new “effective” capacity for each edge e as $\tilde{r}_e = \sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} a_{te} \tilde{x}_{it}$, where \tilde{x}_{it} is the value of the routing solution from phase 1.
2. We define the new per-unit capacitance as $\tilde{C}_e^u = \frac{\tilde{C}_e}{\tilde{r}_e}$, where \tilde{C}_e is the value of the edge capacitance from the solution found in phase 1.

With these definitions, the formulation of phase 2 is the following integer linear program:

$$\min_{x, s, \epsilon} \sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} \alpha_i V_i^2 \left(\sum_{e \ni t} \tilde{C}_e^u \right) x_{it} + \sum_{i=1}^{N_d} M_1 s_i + \sum_{\forall e \in \mathcal{E}} M_2 \epsilon_e \quad (\text{PGR-P2})$$

$$\left\{ \begin{array}{ll} \sum_{t \in \mathcal{T}_i} x_{it} + s_i = 1 & \forall i = 1, \dots, N_d \\ \sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} a_{te} x_{it} \leq \tilde{r}_e + \epsilon_e & \forall e \in \mathcal{E} \\ \sum_{i=1}^{N_d} \sum_{t \in \mathcal{T}_i} w_{it} x_{it} \leq W_0 (1 + \beta) & \\ 0 \leq \epsilon_e \leq r_e - \tilde{r}_e & \forall e \in \mathcal{E} \\ x_{it} = \{0, 1\} & \forall i = 1, \dots, N_d, \forall t \in \mathcal{T}_i \\ s_i \geq 0 & \forall i = 1, \dots, N_d. \end{array} \right.$$

The first term in the objective expression is summation of an estimate of the power of the nets where $(\sum_{e \ni t} \tilde{C}_e^u)$ is the fixed approximate per-unit capacitance of edge e which contains route t and is obtained using the solution of phase 1 as discussed before. The first set of constraints ensures that at most one route is selected per net, otherwise a heavy penalty of M_1 is associated if $s_i \neq 0$, and this is reflected in the second term of the objective function. The second set of constraints enforces the new utilization of each edge to be $\tilde{r}_e + \epsilon_e$, where ϵ_e is a new variable which is heavily penalized by a large factor M_2 in the objective function if $\epsilon_e \neq 0$. In other words, we highly penalize if the rerouting of a net causes a larger edge utilization compared to phase 1. This in effect forces the routing process to keep the *mismatch* in the edge utilizations as small as possible which translates in the capacitance (which is function of utilization) to remain close to phase 1. We also enforce $\epsilon_e + \tilde{r}_e \leq r_e$ to ensure the edge utilization is not beyond its actual capacity r_e in the fourth set of constraints. Finally, the third set of constraints ensures the increase in wirelength is bounded by factor β .

2) *Solving using Price-and-Branch*: The solution procedure is quite similar to the one explained in the previous Section III-B for phase 1. Here, we just note the differences. We denote the restricted master problem by (RMLP-P2) and its solution by $(\hat{x}, \hat{s}, \hat{\epsilon})$. The dual of the restricted master is denoted by (D-RMLP-P2) and its solution is $(\hat{\lambda}, \hat{\pi}, \hat{\theta})$, corresponding to the first, second and third set of inequalities in relaxed (PGR-P2), respectively.

The initial set \mathcal{S}_i is set to *all the candidate routes generated from phase 1*. This helps to quickly generate a high quality solution for phase 2. It also ensures that the solution of phase 1 is included as a *feasible* solution in phase 2.

The pricing condition is given by the following inequality $\hat{\lambda}_i > \alpha_i V_i^2 (\sum_{e \ni t} C_e^u) - \sum_{e \in t} (\hat{\pi}_e + \hat{\theta})$ and is used to define the edge weights given by $w_e = \alpha_i V_i C_e^u - \hat{\pi}_e - \hat{\theta}$, $\forall e \in \mathcal{E}$.

D. Decomposition

To accelerate solving the two-phase formulation, we apply a simple problem decomposition. We recursively divide the chip into a set of rectangular subregions while balancing the total number of nets that fall inside each subregion. We use the initial WL-optimized solution of [5] to guide this process. We stop when the number of nets at each subregion is at most 3000, which we empirically determined for our experimented benchmarks from the ISPD2008 suite [15].

Each subproblem is then defined as one rectangular subregion with the set of nets assigned to it. If a net passes from multiple subregions, we force the terminal location on the subregion boundary to be fixed from the initial WL-optimized solution. This allows independent solving of each subproblem without the hassle of later connecting the segments of a route in adjacent subregions. The subproblems are then (one-time) parallel-solved to get the final solution.

Please note, the main difference between our decomposition procedure and [10] is the use of the initial WL-optimized solution to fix the terminal locations on the subregion boundaries and thus avoid later connecting adjacent subproblems.

IV. SIMULATION RESULTS

A. Benchmark Instances

In order to test our solution procedure and determine whether or not significant power savings were possible without increasing wirelength, we modified known benchmarks to include multi-supply voltages. Modifying the benchmarks required us to generate timing data, power data, and place level converters. We implemented the

procedure of [2] to generate voltage islands for two voltage levels of $V_L = 0.9V$ and $V_H = 1.1V$. The procedure required a sequential netlist with gate-level delay and power models.

Timing Modeling: We assumed the locations of the sequential elements in the ISPD 2008 benchmarks using the following procedure. First, we obtained a Directed Acyclic Graph (DAG) representation of the benchmarks from the variation provided by the ISPD 2006 placement benchmarks [16]. Using the placement benchmarks, we obtained a DAG by starting from the designated Primary Inputs and traversing in forward direction until reaching the Primary Outputs. We also assumed the nets with more than 50 terminals to be clock trees to identify sequential elements.

We then assumed the delay of each cell (or node in the DAG) is proportional to its size (for unit load) where the unit delay was assumed to be of the inverter of the 45nm library [7] used in this work. We considered loading in our cell delay modeling to be proportional to the cell size which was also given in the placement benchmarks.

Power Modeling: We randomly and uniformly generated the activity factors of each net to be between 0.1 and 0.9. The 45nm library used in this work contained information about the total capacitance (area, fringe, coupling) for each of the 8 metal layers. We used the method described in Section III to extract piece-wise linear model for C_e and C_e^u for each of the 8 metal layers. For each metal layer, we considered the minimum wire size given in the library. To map edge utilization to spacing, we assumed the length of each edge of the GR grid to be 2μ ; for a given utilization we assumed the maximum spacing between the routes mapped to the same GR edge.

Level Converter Placement: After voltage island generation, we needed to decide the locations of the level converters (LCs). (The procedure in [2] didn't specify these locations). For simplicity, we inserted the LCs on the initial WL-optimized solution that was taken from [5]. The LCs were inserted for any net that had a source terminal driving one or more sink terminals. The procedure minimized the number of LCs and placed them as close as possible to the sink terminals, subject to the available whitespace. The whitespace inside each global bin was derived by evaluating (both) the placement and GR variations of the ISPD benchmarks.

B. Results

Using the initial WL-optimized solution of [5], and after fixing the locations of LCs, we applied net decomposition (as described in Section 2.1). Table I reports the number of nets, decomposed nets, and LCs in columns 2, 3, 4, respectively. We then applied our power-driven GR procedure using a wirelength degradation factor of $\beta = 0$, so *no* wirelength degradation was allowed. We used CPLEX 12.0 [17] to solve our two-phase formulation, and parallel-processed the subproblems by submitting the jobs to a grid of CPUs of 2GB memory. The number of subproblems (same as number of processors) is given in column 5 (#SP) in Table I.

We then compared three routing solutions.

- The initial WL-optimized solution of [5];
- The solution after applying phase 1, obtained by solving the formulation (PGR-P1);
- The solution by further applying phase 2, obtained by solving (PGR-P1) followed by (PGR-P2).

For each case, we report the wirelength (WL), the total capacitance (C) ($\sum_{i=1}^{N_d} C_i^{route}$, where C_i^{route} is defined in (2)), given in units fF , and the GR power metric P from (1), excluding the constant portions of the expression.

TABLE I
RESULTS FOR ISPD 2008 BENCHMARKS. THE WL IS SCALED TO 10^5 . POWER AND CAP. ARE SCALED TO 10^3 .

Bench	#Net	# Net _d	#LC	# SP	initial solution ([5]+LC)			phase 1			phase1+phase2		
					W ₀	C	P	-WL(%)	-C(%)	-P(%)	-WL(%)	-C(%)	-P(%)
adaptec1	177K	197K	20K	130	54.2	953.3	432.2	0.05	11.70	8.57	0.07	15.48	16.17
adaptec2	208K	224K	17K	195	53.0	750.0	336.9	0.12	10.34	6.93	0.14	14.57	15.13
adaptec3	368K	411K	43K	359	132.7	2187.0	1056.8	0.01	11.51	8.67	0.34	13.55	13.94
adaptec4	401K	437K	36K	296	123.0	1613.8	751.1	0.02	12.16	8.46	0.04	16.92	17.20
adaptec5	548K	632K	85K	454	158.7	2543.0	1199.6	0.38	8.60	6.08	0.43	10.23	10.88
newblue1	271K	287K	16K	195	47.0	612.2	318.9	0.11	13.39	9.87	0.22	17.45	18.40
newblue2	374K	421K	47K	312	77.6	894.9	453.2	0.04	14.19	7.87	0.09	19.20	19.34
newblue4	531K	610K	79K	462	133.7	1955.4	927.7	0.02	13.39	9.61	0.54	17.45	17.61
newblue5	892K	975K	84K	658	234.7	3405.3	1469.9	0.89	11.55	6.75	0.86	14.00	13.47
newblue6	835K	926K	91K	532	180.2	2834.9	1367.0	0.62	12.56	9.35	0.57	16.35	17.80
newblue7	1647K	1719K	72K	670	360.2	5004.4	2201.8	0.01	15.12	11.20	0.17	19.63	20.93
bigblue1	197K	222K	26K	152	57.0	1110.4	619.3	0.23	10.68	7.20	0.16	12.17	12.56
bigblue2	429K	472K	43K	275	92.4	1283.8	560.7	0.14	11.64	7.86	0.10	14.76	14.33
bigblue3	666K	725K	60K	453	133.0	1664.6	815.0	0.91	15.22	10.99	0.93	20.25	20.31
bigblue4	1134K	1184K	51K	509	233.0	3006.6	1254.3	0.18	16.03	12.12	0.28	22.31	22.46
Avg.								0.25	12.54	8.77	0.34	16.29	16.70

The results are reported in Table I in columns 6 to 14. For the initial solution, we report the wirelength (W_0) of the NTHU-R2.0 routes that have been augmented with the extra via-only segment(s) to connect the LC(s) to the original routes. (As a result, there is slight increase in wirelength compared to the numbers reported in the work [5]). For the solutions of phase 1 and phase 2, we report only the percentage *improvement* in WL, C, and P, all with respect to the initial solution.

As can be seen, applying phase 1 of the power-reduction heuristic results in significant saving of 8.77% in P. Recall, the savings are solely due to capacitance reduction (as can be seen from the higher improvement rate in C compared to P). By further applying phase 2, we see additional improvement in P (on average 16.70%). The improvement in C is slightly larger than phase 1, even though phase 1 solely focuses on optimizing C. This is because we start phase 2 by including all the candidate routes generated from phase 1. Notice that in both phase 1 and phase 2 there is improvement (reduction) in WL compared to W_0 . It is important to note that no extra overflow was introduced in the power-optimized solutions.

In our simulations, we explicitly bounded the runtime for phase 1 and phase 2. The wall clock runtime of all benchmarks for phase 1 and phase 2 were set to 30min and 40min, respectively. The number of processors (same as subproblems) is given in column 5.

V. CONCLUSIONS

We proposed a formulation for minimizing an interconnect power metric for global routing for design with multi-supply voltage. Power minimization is after an initial wirelength-optimized solution is obtained. We presented a mathematical formulation which considered power saving opportunities by reducing the area, fringe and congestion-dependent coupling capacitances at each metal layer, while accounting for the activity and supply voltage of each route segment. We showed significant savings in the power metric for global routing without any degradation in wirelength or overflow.

REFERENCES

- [1] R. Shelar and M. Patyra, "Impact of local interconnects on timing and power in a high performance microprocessor," in *ACM ISPD*, 2010, p. invited talk [Online] http://ispd.cc/slides10/8_01.pdf.
- [2] R. L. S. Ching, E. F. Y. Young, K. C. K. Leung, and C. C. N. Chu, "Post-placement voltage island generation," in *ICCAD*, 2006, pp. 641–646.
- [3] L. Guo, Y. Cai, Q. Zhou, and X. Hong, "Logic and layout aware voltage island generation for low power design," in *ASP-DAC*, 2007, pp. 666–671.
- [4] H. Shojaei, T.-H. Wu, A. Davoodi, and T. Basten, "A Pareto-algebraic framework for signal power optimization during global routing," in *ISLPEd*, 2010.
- [5] Y.-J. Chang, Y.-T. Lee, and T.-C. Wang, "NTHU - Route 2.0: A fast and stable global router," in *ICCAD*, 2008, pp. 338–343.
- [6] M. Cho, K. Lu, K. Yuan, and D. Z. Pan, "Boxrouter 2.0: A hybrid and robust global router with layer assignment for routability," *ACM Trans. Design Autom. Electr. Syst.*, vol. 14, no. 2, 2009.
- [7] "Nangate 45 nm open cell library, [online] <http://www.nangate.com>," 2008.
- [8] C. Barnhart, E. L. Johnson, G. L. Nemhauser, M. W. P. Savelsbergh, and P. H. Vance, "Branch-and-price: Column generation for solving huge integer programs," *Operations Research*, vol. 46, pp. 316–329, 1996.
- [9] D. G. Jørgensen and M. Meyling, "A branch-and-price algorithm for switch-box routing," *Networks*, vol. 40, pp. 13–26, 2002.
- [10] T.-H. Wu, A. Davoodi, and J. T. Linderoth, "GRIP: Scalable 3D global routing using integer programming," in *DAC*, 2009, pp. 320–325.
- [11] G. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations Research*, vol. 8, pp. 101–111, 1960.
- [12] E. W. Dijkstra, "A note on two problems in connection with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271.
- [13] M. Pan and C. C. N. Chu, "Fastroute 2.0: A high-quality and efficient global router," in *ASP-DAC*, 2007, pp. 250–255.
- [14] J. A. Roy and I. L. Markov, "High-performance routing at the nanometer scale," *IEEE Trans. on CAD of Integrated Circuits and Systems*, vol. 27, no. 6, pp. 1066–1077, 2008.
- [15] ISPD 2008 global routing contest and benchmark suite [online]<http://www.sigda.org/ispd2008/contests/ispd08rc.html>. [Online]. Available: [online]<http://www.sigda.org/ispd2008/contests/>
- [16] ISPD 2006 placement contest and benchmark suite. [Online]. Available: [online]<http://www.sigda.org/ispd2006/contest.html>
- [17] *Using the CPLEX Callable Library, Version 9*, CPLEX Optimization, Inc., Incline Village, NV, 2005.