Smoothening Max-terms and Analytical Minimization of Half-Perimeter Wirelength*

Andrew A. Kennings and Igor L. Markov

Cypress Semiconductor, 3901 North 1st Street, San Jose, CA 95134

EECS Department, U. Michigan, 1301 Beal Avenue, Ann Arbor, MI 48109-2122

ank@cypress.com, imarkov@eecs.umich.edu

Abstract

This work addresses a class of optimization problems arising in engineering, where the objective function contains non-differentiabilities, and yet needs to be minimized analytically. While large classes of non-differentiabilities can be successfully smoothened by approximations, existing techniques fail to provide a *symmetric*, *smooth* and *computationally convenient* approximation for the multivariate *max* function with provable properties. Our work proposes such an approximation and immediately applies it to a hypergraph placement problem, previously addressed by heuristic transformation of hypergraphs into graphs. Empirical validation is performed by comparing our implementations to several optimal but unacceptably laborious methods, including network flows and linear programming with CPLEX.

Keywords

VLSI placement, wirelength, half-perimeter, force-directed, timing-driven, approximation, linear, non-linear, analytical, large-scale.

*Early results of this work were reported in January 2000 at *Asia and South Pacific Design Automation Conference* in Japan.

1 Introduction

Objective functions optimized in engineering problems often contain non-linear terms of many variables that imply analytical minimization procedures. Yet, those terms may be connected through *absolutevalue* functions, *max* functions or more general *case* functions whose non-differentiabilities break numerical stability of fast analytical methods. For example, dealing with transistor sizing, [20] proposed to minimize $(1/\lambda)\ln(e^{\lambda x_1} + e^{\lambda x_2} + \ldots + e^{\lambda x_k})$ using Lagrangian relaxation instead of the original function $\max(x_1, x_2, \ldots, x_k)$. [23] addressed min-wirelength graph placement by empirically *linearizing* the computationally easy quadratic objective (i.e., minimizing the quadratic objective instead of linear, in an iteration). The linear convergence of that method was shown in [1] which also demonstrated a faster, quadratically convergent Primal-Dual method enabled by substituting $|x_i - x_j|$ with $\sqrt{(x_i - x_j)^2 + \beta}$ for small $\beta > 0$. [3] formalized this β-regularization, proved a series of convenient properties and demonstrated applications to convex signal delay models for VLSI interconnects. Recently, [32] proposed using stochastic penalty functions to achieve a smoothening effect in constrained optimization and proved that under certain assumptions their Primal-Dual algorithm almost surely converges. [4] studied optimization of a lower semi-continuous extended-valued convex function by smoothening.

In this work, we deal with an application where a multi-variate *max* function is the source of problems. The regularization proposed in [20] implies a large number of [expensive] exponent computations, and the β-regularization implies a rather difficult gradient computation. Moreover, we show in this work that a straightforward application of β-regularization leads to a function that is not symmetric in the original variables. Therefore, we propose a new deterministic p, β-regularization and prove its convenient properties. Empirical results are produced by applying our new analytical placement algorithm on circuits from the industry.

Analytical placers are increasingly important in physical design as process technology advances and design complexity increases. The fact that interconnect delays dominate device and cell delays in large netlists entails a global optimization setting that focuses on the area and performance costs of *interconnect* rather than *devices*. A simple mathematical quantity (wirelength estimate) representing interconnect delay is minimized by placing numerous shapeless "modules" that represent individual cells, transistors or other design units. They locate modules (cells or macros) so as to minimize a wirelength estimate representing a cumulative measure

of interconnect delay and utilization; some algorithms also minimize specific timing-critical paths. Placement solutions must satisfy various combinatorial constraints, e.g., use only prescribed module locations, avoid module overlaps etc. These constraints are temporarily relaxed for analytical placement and are later taken care of by specialized code which may itself make repeated calls to analytical placement. For example, a placement with excessive module overlaps or overutilization of routing resources can be spread out using *min-cut partitioning* [26, 25, 12], *transportation* [29] or *force-directed techniques* [6].

In top-down placement, the layout area is recursively split into blocks; cells inside each block are dealt with under the assumption that all other cells are fixed. Analytical placers do not handle well large macros with non-trivial geometries [30] and are difficult to apply when insufficiently many terminals lead to numerical degeneracy. However, analytical placers may be helpful before or instead of min-cut partitioning steps at middle and lower levels of top-down placement when terminal counts increase due to terminal propagations from other placement blocks and if timing constraints are considered. Analytical placers are also useful for quick delay budgeting before non-overlapping and routable placements are available [22]. Therefore, min-cut partitioning and analytical placement are not direct competitors, but rather complement each other when timing considerations are important.

Analytical placers typically transform a circuit hypergraph into a graph prior to solving any optimization problems. Each hyperedge is modeled by a *star* [23, 21] or a *clique* [10, 27, 26] of edges. Early algorithms [27, 26] used squared wirelength objectives since global module placements required the solution to a single system of equations. However, the squared wirelength objective tends to overemphasize the minimization of long wires at the expense of short wires; this increases the demand on routing resources, thereby leading to a poorer layout [16]. Recent analytical placers [23, 1] rely on a linear wirelength objective that is optimized, e.g., by iterated approximations with quadratic wirelength objectives [23].

Placement qualities are typically evaluated using the half-perimeter wirelength (HPWL) of the circuit hyperedges, however this objective cannot be directly pursued after nets are converted into stars or cliques. In this paper, we seek to eliminate the divergence from the minimization of HPWL in analytical placers.

HPWL minimization is possible via linear programming (LP) [9, 31], but is too computationally intensive due to the sizes of the resulting linear programs. Moreover, inclusion of non-linear terms is not straightforward. [22]

addresses convex delay terms and describes a general *linearization* procedure that constructs piece-wise linear approximations, but finds it is too costly to solve directly as the complexity of linearizations increases with required precision.[1] An optimal algorithm for HPWL minimization based on classic max-flow computation was proposed by Hur and Lillis [8]. This technique, however, makes the inclusion of non-linear timing constraints [3] even harder. We thus seek a convex nonlinear approximation to enable easy inclusion of delay terms into analytical placement formulations.

In this work we

- present a new [smooth] regularization of the multi-variate max function
- prove its mathematical properties, such as convergence to the original function and error bounds
- apply the new regularization to the half-perimeter wirelength function
- propose the first analytical algorithm for half-perimeter wirelength minimization that bypasses traditional graph models of multi-pin nets and is naturally amenable to non-linear timing terms [3]. Thus the expensive *linearization* of delay in [22] can be avoided.
- empirically confirm our approach by developing a "proof-of-concept" implementation that found solutions within 12% of optimum and outperformed graph-based wirelength minimizations.

Our proposed nonlinear and convex approximation to HPWL is based on the observation that HPWL of multi-pin nets is a convex, but not everywhere differentiable function with singularities arising from "max" functions. Based on this observation, we extend the recently proposed *function smoothing techniques* in [3] for various VLSI layout problems to make them applicable to HPWL.

Section 2 demonstrates the difficulty of the direct minimization of HPWL due to its non-differentiability and reviews previous work. Section 3 proposed regularizations that approximate HPWL with arbitrarily small relative error and enable graph-free HPWL minimization via unconstrained smooth and convex optimization. Section 4 describes our analytical placer and presents numerical results for industrial testcases in the context

[1] Their faster algorithm for a particular delay budgeting problem is rather difficult to implement as it relies on min-cost flows and graph-based simplex methods.

of top-down placement. Conclusions are given in Section 5. Appendix A proves lemmas about wirelength computations used in the text.

2 Review of analytical placement

Circuits are represented by weighted hypergraphs $G_H(V_H, E_H)$ with vertices $V_H = \{v_1, v_2, \cdots, v_n\}$ corresponding to modules, and hyperedges $E_H = \{e_1, e_2, \cdots, e_m\}$ corresponding to signal nets. Vertex weights correspond to module areas, while hyperedge weights correspond to criticalities and/or multiplicities. Vertices are either *fixed* or *free*. Hyperedge $e_k \in E$ connects $p_k \geq 2$ vertices and each vertex $v_i \in V$ is incident to $d_i \geq 0$ hyperedges. p_k and d_i are respectively called the vertex and hyperedge *degrees*, and are typically very small. We say that v_i has d_i pins, and e_k has p_k pins, for a total of $P = \sum_{k=1}^{m} d_k = \sum_{i=1}^{n} p_i$ pins in the hypergraph. Module placements in x and y directions are captured by the *placement vectors* $\mathbf{x} = (x_1, \ldots, y_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$.

2.1 Hypergraph placement

Let C_k be the index set of hypergraph vertices incident to net $e_k \in E_H$. The x-direction HPWL estimate is given by

$$HPWL_x(\mathbf{x}) = \sum_{e_n \in E_H} \max_{i,j \in C_k} |x_i - x_j|. \tag{1}$$

HPWL is a *convex* function of $\mathbf{x}$ since $|x_i - x_j|$ is convex for all i, j. However, HPWL is not *strictly convex* and most often has uncountably many minimizers. HPWL minimization requires fixed vertices, with at least two different locations (otherwise placing all vertices to the same location will achieve an optimal solution with wirelength 0). Fixed vertices in circuit hypergraph are provided by I/O pads and external pins. The non-differentiability of the max function disables classic smooth minimization techniques such as Newton method. It can be shown that any single iteration of *the steepest descent* will end up in a solution where gradient is not well-defined and the new steepest descent direction is not easily available.[2]

In [31], the HPWL estimate is converted into an equivalent linear program (LP) by adding, for each net e_k,

[2] An even bigger problem is demonstrated by a 3-clique of free vertices that is connected to a fixed vertex by one edge. As soon as all free vertices are located at the same point, no movement of a single vertex can improve wirelength, while moving all three toward the fixed vertex will lead to the optimal placement.

upper and lower bound variables U_k and L_k. The cost of the net is measured as the difference between the two, and in an optimal solution each U_k and L_k will be equal to the rightmost and leftmost module locations of net e_k. Each variable comes with p_i inequality constraints that restrict $U_j(L_j)$ to be larger (smaller) than the locations of every module incident to the net. Thus, n nets and m modules are represented by $m + 2n$ variables and $2P$ constraints. While LP returns optimal placements, the large instance sizes effectively preclude any application to fast placement of large hypergraphs.

2.2 Reduction to graphs

Circuit hypergraphs are typically transformed into graphs in which each hyperedge is represented by a group of equally weighted edges. The *unoriented star model* adds a new *center* vertex and represents the original net by edges connecting the center to previously existing vertices (modules) [23, 21]. The *clique model* (e.g., [10, 13]) connects all pairs of vertices (modules) incident to the original hyperedge by edges of non-unit weight.[3] Clique models of large hyperedges become prohibitively expensive due to the quadratic edge count. Therefore, large hyperedges are typically modeled by stars or dropped completely.

Wirelength estimates for individual edges of a graph are weighted and added up to produce a total wirelength estimate. For an edge that connects modules with abscissae x_1 and x_2, the most popular x-wirelength estimates are (a) *linear* (Manhattan) $|x_1 - x_2|$ and (b) *square* (Euclidean) $(x_1 - x_2)^2$; the y-wirelength is computed in the same way and added to the x-wirelength. While both functions are *convex*, only the square wirelength is *strictly convex* when all graph vertices are are reachable from fixed vertices, which guarantees a unique minimizer.

Minimization of squared (quadratic) wirelength

$$\min_{\mathbf{x}} \{ \textstyle\sum_{i>j} a_{ij}(x_i - x_j)^2 : \mathbf{Hx} = \mathbf{b} \} \tag{2}$$

($\mathbf{H}$ represents various linear constraints) is the easiest because (see [27, 26]) the unique minimizer is obtained by solving a single system of linear equations, either positive-definite or symmetric-indefinite depending on

[3] A shortcut [23] computes the squared wirelength of a p-clique with uniform edge weights $\frac{2}{p}$ as the squared wirelength of a star whose center vertex is placed at the center of gravity of its p vertices. In contrast to the general star model, the center of the star is not an independently placed vertex.

the approach. While good public-domain implementations of linear system solvers are available, the squared wirelength objective tends to provide lower-quality placements; a comparison by Mahmoud et al. [16] concludes that the linear wirelength objective is superior.

Linear wirelength minimization also relies on the above reduction of circuit hypergraph to a graph

$$\min_{\mathbf{x}}\{\textstyle\sum_{i>j} a_{ij}|x_i - x_j| : \mathbf{Hx} = \mathbf{b}\} \tag{3}$$

Being neither differentiable nor strictly convex, it is not amenable to Newton-type methods. GORDIAN-L [23] minimization heuristic uses iterated quadratic minimizations

$$\min_{\mathbf{x}^\nu}\{\textstyle\sum_{i>j} \frac{a_{ij}}{|x_i^{\nu-1} - x_j^{\nu-1}|}(x_i^\nu - x_j^\nu)^2 : \mathbf{Hx}^\nu = \mathbf{b}\} \tag{4}$$

where $\mathbf{x}^{\nu-1}$ and $\mathbf{x}^\nu$ denote the vectors of vertex positions at iterations $\nu-1$ and ν. A quadratic objective is used to avoid the non-differentiability of the objective of (3), but the coefficients a_{ij} are updated at each iteration to approximate the linear wirelength. As an alternative, the *regularization* of (3)

$$\min_{\mathbf{x}}\{\textstyle\sum_{i>j} a_{ij}\sqrt{(x_i - x_j)^2 + \beta} : \mathbf{Hx} = \mathbf{b}\} \tag{5}$$

was proposed in [1] with two solution methodologies: a linearly-convergent fixed point method and a novel primal-dual Newton method with quadratic convergence. Testing in [1] illustrated tradeoffs in values of $\beta > 0$ versus time and difficulty. The interpretation of the GORDIAN-L heuristic as a special case of $\beta = 0$ of a fixed point method was also provided. The techniques of **β-regularization** have been further extended in [3].

3 Multivariate regularization for HPWL

Motivated by the convexity of HPWL, we seek a technique for its minimization via smooth convex Newton-type methods without graph approximations.

For two-pin hyperedges, the β-regularization [1, 3] can be used to approximate edgelength by a smooth

convex function; such an approximation overestimates the original function by at most $\sqrt{\beta}$ and used in (5) to approximate graph edgelengths. In order to apply it to hyperedges of degree ≥ 3 and the HPWL, one can nest maximum functions, e.g., for a 3-pin net, $\max\{|x_1 - x_2|, |x_1 - x_3|, |x_2 - x_3|\}$ can be rewritten as $\max\{|x_1 - x_2|, \max\{|x_1 - x_3|, \max\{|x_2 - x_3|\}\}\}$, and recursively apply the following regularization:

Example [3]: The function $\max\{a, b\} = \frac{a+b+|a-b|}{2}$ can be regularized with $\frac{a+b+\sqrt{|a-b|^2+\beta}}{2}$, and $\min\{a, b\} = \frac{a+b-|a-b|}{2}$ with $\frac{a+b-\sqrt{|a-b|^2+\beta}}{2}$.

Unfortunately, the derivatives of the resulting regularization are difficult to compute. Moreover, the regularization is not symmetric and will send line search in wrong search directions.

A simple and computationally efficient alternative is available. **First**, note that we only consider max{} of non-negative numbers, **second**, $\| (a_1, a_2, \ldots, a_k) \|_\infty = \max\{|a_1|, |a_2|, \ldots, |a_k|\}$, **third**, the L_p-norms $\| (a_1, a_2, \ldots, a_k) \|_p = (\sum_{j=1}^{k} |a_j|^p)^{1/p}$ converge to the L_∞-norm $\| \cdot \|_\infty$ as $p \to \infty$. This is illustrated in Figure 1, where unit-norm curves ("unit circles") for L_p-norms on the plane are seen converging to the the "unit circle" $\| (a_1, a_2) \|_\infty = \max(a_1, a_2) = 1$.

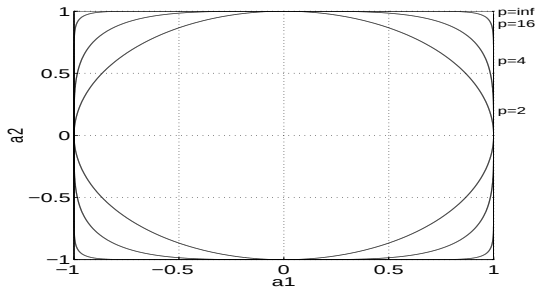

Figure 1: Unit-norm curves for L_p-norms.

Fact 1 For $p > 1$ and $\mathbf{a} = (a_1, a_2, \ldots, a_k)$

(a) $\| \mathbf{a} \|_\infty \leq \| \mathbf{a} \|_{p+1} \leq \| \mathbf{a} \|_p \leq k^{1/p} \| \mathbf{a} \|_\infty$

(b) $\varrho(\mathbf{a}) = \| \mathbf{a} \|_p$ is strictly convex and

infinitely differentiable except at $\mathbf{a} = \mathbf{0}$

Proof We first establish the inequalities in **(a)**.

$\max\{|a_1|, \cdots, |a_k|\} = (\max\{|a_1|^p, \cdots, |a_k|^p\})^{1/p} \leq (|a_1|^p + \cdots + |a_k|^p)^{1/p}$ implies $\| \mathbf{a} \|_\infty \leq \| \mathbf{a} \|_p$ for any $p \geq 1$.

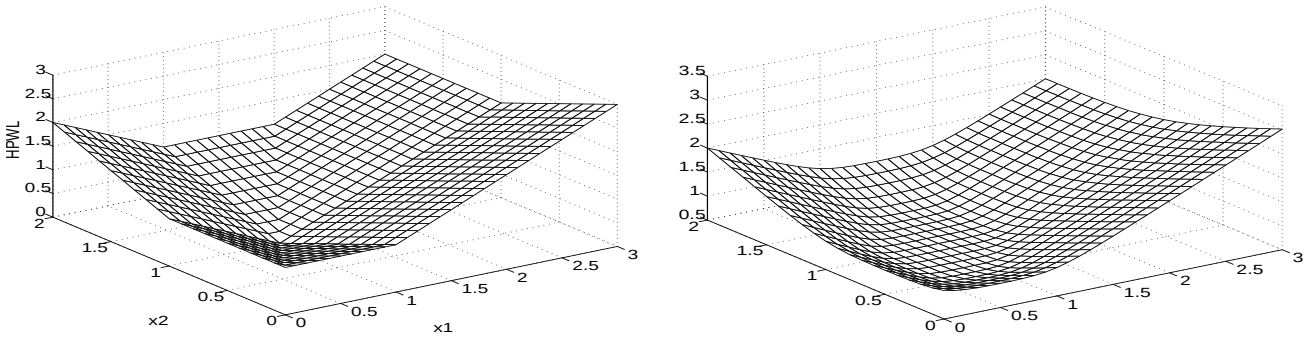

Figure 2: HPWL(left) for a 3-pin net — $\max\{|x_1-x_2|, |x_1-1|, |x_2-1|\}$ over $(x_1, x_2) \in [0,3]\times[0,2]$. The p-regularization (right) is $(|x_1-x_2|^p + |x_1-1|^p + |x_2-1|^p + \beta)^{1/p}$. Here $p = 8$ and $\beta = 1.0e6$.

$\|\mathbf{a}\|_p \geq \|\mathbf{a}\|_{p+1}$ can be shown by induction from the case $k = 2$. Assume positive a_i, x and y. The induction step can now be accomplished by introducing $d_1 = (a_2^p + \ldots + a_k^p)^{1/p}$ and $d_2 = (a_2^{p+1} + \ldots + a_k^{p+1})^{1/(p+1)}$ for which $d_1 \geq d_2$ holds by induction hypothesis. Then $(a_1^p + a_2^p + \ldots + a_k^p)^{1/p} \geq (a_1^p + d_1^p)^{1/p} \geq (a_1^p + d_2^p)^{1/p} \geq (a_1^{p+1} + d_2^{p+1})^{1/(p+1)} \geq (a_1^{p+1} + a_2^{p+1} + \ldots + a_k^{p+1})^{1/(p+1)}$, the middle inequality being case $k = 2$: $(x^p + y^p)^{1/p} \geq (x^{p+1} + y^{p+1})^{1/(p+1)}$. Equivalently $(x^p + y^p)^{p+1} = (x^p + y^p)(x^p + y^p)^p \geq (x^{p+1} + y^{p+1})^p$, where both sides can be interpreted as binomial with equal number of terms. It now suffices to prove that the inequality holds term by term $(x^p + y^p)(x^p)^i(y^p)^{p-i} \geq (x^{p+1})^i(y^{p+1})^{p-i}$ (equal constants canceled out), which follows from $(x^p + y^p) \geq \max\{x^p, y^p\} \geq x^i y^{p-i}$. This concludes the proof of **(a)**.

The differentiability in **(b)** can be proven by taking partials, e.g., $\frac{\partial f}{\partial a_i} = p a_i^{p-1}(\sum_{j=1}^k |a_j|^p)^{\frac{1}{p}-1}$. All n-th partials will be polynomials of $a_i, i = 1..k$ and $(\sum_{j=1}^k |a_j|^p)^{\frac{1}{p}-l}, l = 1..n$. The latter have a pole at $\mathbf{a} = 0$ and are differentiable elsewhere since $\frac{1}{p} - l < 0$. *Strict* convexity in **(b)** can be proven by comparing the Jacobian to zero. □

To approximate the HPWL of an m-pin net, we enumerate all $\frac{m(m-1)}{2}$ pairwise distances of the form $x_i - x_j$ and rewrite the HPWL as their L_∞-norm (see Equation 1), then approximate with L_p-norms.[4]

The multiplicative upper bound of $(\frac{m(m-1)}{2})^{1/p}$ on the overestimation for HPWL of one m-pin net, as given by Fact 1(**a**), appears very loose since all $\frac{m(m-1)}{2}$ distances cannot be equal unless being 0. Tighter bounds can be derived given that the L_∞-norm is applied only to vectors, whose coordinates are all pairwise distances between m points on the real line. Such tighter bounds for our regularization of the HPWL of small nets are

[4]Nets that entail prohibitively many $(\frac{m(m-1)}{2})$ terms can be handled via the L_p-norm taken over edges of a star model with the center located at the center of gravity of the net's pins. See Appendix A for a more detailed explanation.

derived in Appendix A, several numerical values are given in Table 1.[5]

Net size	Upper bounds		Maximal overestimation			
m	loose	tight	p=8	p=16	p=32	p=64
3	$3^{1/p}$	$2^{1/p}$	9%	4%	2%	1%
4	$6^{1/p}$	$4^{1/p}$	19%	9%	4%	2%
5	$10^{1/p}$	$6^{1/p}$	25%	12%	6%	3%
6	$15^{1/p}$	$9^{1/p}$	32%	15%	7%	3%

Table 1: Single net HPWL overestimation by p-regularization.

As follows from Fact 1, overestimating $\max\{|a_1|, |a_2|, \cdots, |a_k|\}$ by its p-regularization $(|a_1|^p + |a_2|^p + \cdots + |a_k|^p)^{1/p}$ removes all nondifferentiabilities except for $\mathbf{a} = \mathbf{0}$. Additional overestimation by β-regularization $(|a_1|^p + |a_2|^p + \cdots + |a_k|^p + \beta)^{1/p}$ smoothens the function at $\mathbf{a} = \mathbf{0}$.

The resulting approximation of HPWL

$$HPWL(\mathbf{x}) \leq HPWL_{reg}(\mathbf{x}) = \sum_{e_n \in E_H} (\sum_{i,j}^{|C_n|} |x_i - x_j|^p + \beta)^{1/p} \tag{6}$$

is a smooth and strictly convex[6] upper bound on exact HPWL with arbitrary small relative error of approximation as $p \to \infty$ and $\beta \to 0$. Figure 2 illustrates the combined p- and β-regularization for HPWL. We note that restricting p to powers of 2 allows for particularly effective computations.

Our regularization subsumes the one in [3] with an important addition of $p \to \infty$ (p was defined differently in [3] and set to 2 for all applications). We set $\beta = (\beta_0 M)^p$ (cf. [3]) where M is the maximal distance between fixed terminals and β_0 is instance-independent.

4 Experimental validation

To compare to optimal solutions found via linear programming, our tests do not include *vertex spreading*, such as equality constraints in GORDIAN/GORDIAN-L or top-down framework in PROUD. We expect that improvements in the fundamental analytical engines translate to complete algorithms.

[5] E.g., for a 3-pin net, the L_p-norm is $(|x_1 - x_2|^p + |x_1 - x_3|^p + |x_2 - x_3|^p)^{1/p}$. Clearly, the three terms cannot be equal (unless 0). Assume an arbitrary ordering $x_1 \leq x_2 \leq x_3$ and maximize the L_p-norm for fixed x_1 and x_3: the maximal overestimation $2^{1/p}$ is reached when x_2 is placed on top of either x_1 or x_3. A similar argument for 4-pin nets yields a tight bound of $4^{1/p}$ even though there are 6 terms involved in the L_p-norm for a 4-pin net.

[6] Strict convexity requires that all free vertices be reachable from fixed vertices. Otherwise there will be multiple optimal solutions, contradicting strict convexity.

For our proposed HPWL minimization, both the objective and the gradient can be computed analytically, but the Hessian computations are hard and time-consuming. Given the crucial nature of second order information, we have implemented the limited memory quasi-Newton method in [14, 18] which uses limited memory BFGS updates to approximate the Hessian.[7] Our implementation maintains storage for seven past iterations. Iterations continue until (i) a prescribed iteration limit (100) OR (ii) the improvement in the objective function is below a prescribed threshold OR (iii) the gradient norm falls below a prescribed threshold. We use line search developed by Jorge J. Moré and David J. Thuente for the MINPACK project in 1983. Although not as accurate as a binary convex line search, the Moré-Thuente is several times faster and has a better cost-performance ratio.

We use five testcases from industry (see Table 2) that are either original ("top-level") placement instances or have arisen on further levels of top-down placement.[8] Test3 through Test5 correspond to placement blocks at various levels in top-down placement of industrial designs of sizes up to 69K cells. Applying analytical placement on higher levels is not practical as insufficient number of fixed vertices leads to degeneracy of the analytical placement model in Section 2. We use the proposed approximation of HPWL in an algorithm called

Instance	Modules		Nets	Design
	Fixed	Free		size
test1	76	200	242	276
test2	545	2686	2840	3.2K
test3	2155	6739	7330	12K
test4	2191	3205	3835	12K
test5	6545	17380	20902	69K

Table 2: Testcase parameters. Test3,4,5 are top-down placement blocks sized at 1/2, 1/4 and 1/4 of their complete designs.

BoxPlace, which we implemented in C++ using the SunPro CC4.2 compiler (`-O5`) and Solaris 2.6 operating system. During experiments, parameters were set $p = 16$ and $\beta_0 = 0.01$.

Table 3 shows that BoxPlace produces better placements than graph-based algorithms and is faster. Its global convergence was tested by running from multiple random initial starting points. Alternatively, we started with a "one-point" solution placing all vertices into one location. This yielded much better *initial* wirelength and

[7] Notably, the BFGS (Broyden-Fletcher-Goldfarb-Shanno) updates are closely related to the DFS (Davidon-Fletcher-Powell) updates [7], except for a slightly different inverse Hessian approximation. BFGS updates typically perform better and are preferred for practical applications.

[8] Several test circuits had disconnected cells not reachable from fixed terminals. To avoid degeneracy and subsequent breakdown of numerical solvers, we assured that only free vertices reachable from fixed vertices have been passed to the solver. Others have been placed in the center of the layout to minimize WL.

faster convergence to a solution comparable to those achieved in the randomized experiment.

HEURISTIC PLACEMENT ALGORITHMS FOR HALF-PERIMETER WIRELENGTH OBJECTIVE											
	BoxPlace from random			BoxPlace from "one-point"				Quadratic		Weiszfeld[1]	
	initial	final WL	CPU◷	initial	final WL	+greed x2	CPU◷	WL	CPU◷	WL	CPU◷
test1	5.73**e7**	6.38e6	0.4	6.72e6	6.39e6	**6.23e6**	0.2	7.66e6	0.11	6.72e6	0.15
test2	3.08**e8**	3.25e7	24.4	3.51e7	3.11e7	**3.01e7**	15.0	4.20e7	3.75	3.51e7	7.12
test3	6.27e6	2.36e6	28.5	2.69e6	2.32e6	**2.21e6**	13.8	2.75e6	19.9	2.67e6	22.4
test4	5.37e6	1.49e6	13.1	1.91e6	1.44e6	**1.38e6**	5.8	1.61e6	5.4	1.78e6	12.9
test5	3.75e7	2.56e7	45.1	1.50e7	1.40e7	**1.33e7**	33.8	1.50e7	72.1	1.47e7	87.6

Table 3: BoxPlace algorithm compared to graph-based algorithms by total HPWL (the sum of x- and y-values). Run times are in seconds on a `Sun Ultra-10/300` MHz and averaged over $x-$ and $y-$ to represent expected runtime in top-down placement.

OPTIMAL PLACEMENT ALGORITHMS WITH HALF-PERIMETER WIRELENGTH									
	Linear Program			Optimal	`LPSolve2.3w`	`CPLEX 6.5.1` CPU◷			Hur-Lillis [8]
	rows	cols	non0s	WL$(x+y)$	CPU◷	`primopt`	`netopt`	`tranopt`	CPU◷
test1	1050	512	2250	5.93e6	1.2	0.54	0.34	0.32	X
test2	15602	6602	32500	2.75e7	6 min	31	9.11	10.74	X
test3	21276	9046	45563	1.98e6	8 hr	30 min	2 min	2.2 min	6.58
test4	43486	17598	93860	1.24e6	1 hr	72	18.44	21.97	48.82
test5	123648	47438	265750	1.19e7	>3days	12 hr	12 min	15 min	3.3min

Table 4: Optimal hypergraph placement implementations for the HPWL objective. Run times are averaged over the $x-$ and $y-$ directions and given in seconds unless indicated otherwise. `LPSolve3.2w` runs were performed on a Sun `Ultra-10/300MHz`, `CPLEX 6.5.1` — on an `IBM RS/6000 3CT` workstation, which measured $1.6 - 1.15$ times slower than the Sun `Ultra-10`. The Hur-Lillis algorithm [8] ran on a `Sun Ultra-1/200MHz` that measured 1.4 slower than the `Ultra-10`. Optimal costs are sums of $x-$ and $y-$ components. `test1` and `test2` are only available in LEF/DEF format, thus we could not run Hur-Lillis on them. $x-$ and $y-$ linear programs have the same numbers of rows and columns, the numbers of nonzeros are averaged.

Table 4 gives optimal placement costs and run times required to achieve those using `CPLEX 6.5.1` and academic implementations: `LPSolve 2.3` with enhancements by David Warme and a network-flow-based optimal algorithm by Hur and Lillis [8].[9] Optimal costs provide a baseline for comparing heuristics. We observe that on our instances BoxPlace achieves placement quality within 12% of optimal. This is, of course, not a guarantee that the same results will be achieved by any other implementation on any other instance, but rather a "proof of concept". In general, the results depend on instance structure (entailed by applications), the values of p and β (greater p and smaller β tend to give better solutions, but require more iterations and longer runtime) and convergence criteria (more strict convergence criteria imply more iterations, thus longer runtime and better

[9]Compared to [31], our linear programs have 50% less additional variables for 2- and 3-pin nets as well as correspondingly fewer constraints for 2-pin nets.

solutions). Clearly, those parameters influence the trade-off between solution quality and runtime, and need to be tuned for particular applications. In our experiments, BoxPlace runs substantially faster than LP-based implementations and the Hur-Lillis algorithm *on larger instances*. We also ran a multi-level FM (MLFM) partitioner on the same circuits, and one start was at least five times faster than BoxPlace. Unlike MLFM, our solver can naturally accommodate additional non-linear terms in the objective function as long as they are convex and differentiable (or can be regularized).

5 Conclusions

We presented a new, symmetric and efficiently computable regularization ("smoothening") of the multi-variate *max*-function and proved its important mathematical properties. As an application, we proposed a fast analytical placement algorithm based on a new approximation of half-perimeter wirelength. Such an approximation can be arranged to have arbitrarily small error, thus trading solution quality for numerical stability during optimization.

Our analytical placement algorithm is the first such to minimize half-perimeter wirelength bypassing traditional net models. Unlike previously known heuristics, it can accommodate convex non-linear delay terms and produces solutions within 12% of optimum. The advantage of our approach compared to [22] where delay terms are linearized to apply linear programming, is that the complexity of non-linear approximations does not grow when better precision is needed.

We implemented a Newton-type optimization algorithm.[10] To avoid the difficulties of computing Hessian information in the context of HPWL minimization, we have adapted a known limited memory quasi-Newton method [15, 14] which *implicitly* keeps track of second order information derived from gradient computations. Comparisons to the authors' implementation of Weiszfeld algorithm [1] confirms the superiority of quadratically convergent methods and goes well with comments in [30] that GORDIAN-L is rather slow. Our techniques are applicable to other objectives, e.g., (trivially) to the piece-wise linear objective in [28].

Among the limitations of our work is the lack of integrated techniques to remove cell overlaps and attention

[10] Recall that [11] pointed out that "second-order" information is important for handling the clique/clustered nature of circuit hypergraphs, and encouraged the use of Newton-type methods in conjunction with twice-differentiable approximations.

to routability. However, these two topic were addressed in recent works on multilevel optimization for large-scale circuit placement [5] and force-directed macro cell placement [17]. The techniques proposed in [5, 17] are compatible with methods advocated in this paper.

Acknowledgements

We thank Prof. Andrew Kahng at UCLA for motivating this work and providing extensive comments, Sung Hur and Prof. John Lillis at UIC for benchmarking their algorithm for us. Professors David Morton and Ross Baldick at UT Austin helped us with CPLEX.

Appendix A.

In this appendix we first show the equivalence between the clique and the star model for quadratic wirelength, which explains our use of the star-model approximation of p-degree wirelength for large nets. Then, we deduce bounds for approximation error of p-regularization (error bounds for β-regularization were proven in [3] and can be trivially added to produce error bounds for p, β-regularization).

Consider real numbers $\{x_i\}_{i=1}^k$ and an integer $p \geq 1$. Define $\mathcal{S}_k^p = \sum_{i>j} |x_i - x_j|^p$. The following equivalence of clique and star models has been pointed out in [23] without proof

Fact 2 Let $x_c = \frac{1}{k}\sum_i x_i$, then $\mathcal{S}_k^2 = \sum_{i>j}(x_i - x_j)^2 = k\sum_i (x_i - x_c)^2$

Proof

$$k\sum_i (x_i - x_c)^2 = k\sum_i x_i^2 + (\sum_i x_i)^2 - 2\sum_i x_i(\sum_j x_j)$$

$$= (k+1)\sum_i x_i^2 + 2\sum_{i>j} x_i x_j - 2\sum_i x_i^2 - 4\sum_{i>j} x_i x_j = (k-1)\sum_i x_i^2 - 2\sum_{i>j} x_i x_j = \sum_{i>j}(x_i - x_j)^2$$

□

Now define $\varepsilon_k^p = \frac{(\mathcal{S}_k^p)^{1/p}}{\max_{i,j}\{|x_i - x_j|\}}$. Obviously, $\varepsilon_2^p = 1$. Fact 1(a) implies that $\varepsilon_k^p \geq 1$ for $p \geq 3$.

Lemma 3 For even p, $\varepsilon_k^p \geq (1 + \frac{2(k-2)}{2^p})^{1/p}$

Proof Assume sorted order $x_i \leq x_{i+1}$, fix x_1 and x_k. Minimize ε_k^p as a function of $x_i, i = 2..k-1$ by minimizing

$\mathcal{S}_k^p$. The latter is a smooth strictly convex function, whose unique minimizer zeroes its gradient

$$\frac{\partial \mathcal{S}_k^p}{\partial x_i} = \sum_{i \neq j} p(x_i - x_j)^{p-1}$$

It can be seen that setting all $x_i, i = 2..k-1$ to $\frac{x_1+x_k}{2}$ zeroes all partials. Indeed, for any $i = 2..k-1$, the summation above has only two nonzero terms $p(x_i - x_1)^{p-1} + p(x_i - x_k)^{p-1}$. However, since $x_i - x_1 = -(x_i - x_k)$ and $p-1$ is odd, they cancel out.

For this minimizer $\mathcal{S}_k^p$ evaluates to $|x_k - x_1|^p + 2(k-2)(\frac{|x_k - x_1|}{2})^p$ and $\max_{i,j}\{|x_i - x_j|\} = |x_k - x_1|$. □

Lemma 4 $\varepsilon_k^p \leq (\lfloor \frac{k}{2} \rfloor \lceil \frac{k}{2} \rceil)^{1/p}$

Proof Not unlike in the previous proof, only need to maximize the strictly convex function $\mathcal{S}_k^p$ by varying all k arguments in a segment between the fixed maximum and minimum. In other words, the function is defined over a k-dimensional cube. The convexity and compactness of the domain and the strict convexity of the function imply that its local maxima are necessarily reached at the boundary. Successively applying this fact to the faces of the k-dimensional cube that make up the boundary, we conclude that all local maxima must be reached in the "corners" of the cube. In other words, each of k $\{x_i\}$ is either at the minimum or at the maximum. Since clique vertices are now indistinguishable, $\mathcal{S}_k^p$ only depends on the number of vertices assigned to the minimal value t. Namely, $\mathcal{S}_k^p = t(k-t)(\max_{i,j} |x_i - x_j|)^p$ and is maximized when $2t = k$. For odd k the best integer t are $\lfloor \frac{k}{2} \rfloor$ and $\lceil \frac{k}{2} \rceil$. □

References

[1] C. J. Alpert, T. F. Chan, D. J. H. Huang, A. B. Kahng, I. L. Markov, P. Mulet and K. Yan, "Faster Minimization of Linear Wire Length for Global Placement", *Proc. ISPD '97*, pp. 4-11.

[2] C. J. Alpert, A. E. Caldwell, T. Chan, D. J.-H. Huang, A. B. Kahng, I. L. Markov and M. Moroz, "Analytical Engines Are Unnecessary in Top-down Partitioning-Based Placement" *VLSI Design* (1999), to appear.

[3] R. Baldick, A. Kahng, A. Kennings and I. Markov, "Function Smoothing with Applications to VLSI Layout", *Proc. ASP-DAC '99*, pp. 225-228.

[4] J. R. Birce, L. Qi, Z. We, "Convergence analysis of some methods for minimizing a nonsmooth convex function". *Journal of Optimization Theory and Applications*, vol.97, (no.2), Plenum, May 1998. p.357-83.

[5] T. F. Chan, J. Cong, T. Kong and J. R. Shinnerl, "Multilevel Optimization for Large-Scale Placement", *In Proc. Intl. Conf. Computer-Aided Design 2000*, pp. 171-176.

[6] H. Eisenmann, F. M. Johannes, "Generic Global Placement and Floorplanning". *Proc. DAC '98*, pp. 269-274.

[7] R. Fletcher and M.J.D. Powell, "A Rapidly Convergent Descent Method for Minimization", Computer J. 6, 1963, pp. 163-168.

[8] S.W. Hur and J. Lillis, "Relaxation and Clustering in a Local Search Framework: Application to Linear Placement", *Proc. DAC '99*, pp. 360-366.

[9] M. A. B. Jackson and E. S. Kuh, "Performance-Driven Placement of Cell Based IC's", *Proc. DAC '89*, pp. 370-375.

[10] M. Hanan, P. K. Wolff, and B. J. Agule, "A Study of Placement Techniques." *J. Design Automation and Fault-Tolerant Computing*, vol. 2, 1978, pp. 28–61.

[11] A. Kennings and I. L. Markov, "Analytical Minimization of Half-perimeter Wirelength" *Proc. ASP-DAC 2000*, pp. 179-184.

[12] T. Koide at al, "Par-POPINS: a Timing-driven Parallel Placement Method With the Elmore Delay Model For Row Based VLSIs", *Proc. ASP-DAC '97*, 1997. pp. 133-40.

[13] T. Lengauer , *Combinatorial Algorithms for Integrated Circuit Layout*, Wiley-Teubner, 1990.

[14] D. C. Liu and J. Nocedal, "On the Limited Memory BFGS Method For Large Scale Optimization", *Mathematical Programming* 45 (1989), pp. 503-528.

[15] D. G. Luenberger, "Linear and Nonlinear Programming", 2nd Edition, Addison Wesley, 1984.

[16] I. I. Mahmoud, K. Asakura, T. Nishibu and T. Ohtsuki, "Experimental Appraisal of Linear and Quadratic Objective Functions Effect on Force Directed Method for Analog Placement", *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences* 4(E77-A), 1994, pp. 710-725.

[17] F. Mo, A. Tabbara and R. K. Brayton, "A Force-Directed Macro-Cell Placer", *In Proc. Intl. Conf. on Computer-Aided Design 2000*, pp. 177-180.

[18] J. Nocedal, "Large Scale Unconstrained Optimization", *The State of the Art in Numerical Analysis*, Ed. A Watson and I. Duff, Oxford University Press, 1996.

[19] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[20] A. E. Ruehli, P. K. Wolff, Sr. and G. Goetzel, "Analytical Power Timing Optimization Techniques For Digital Systems", in *Proc. of the 14th ACM/IEEE Design Automation Conference*, pages 142-146, 1977.

[21] B. M. Riess and G. G. Ettelt, "Speed: Fast and Efficient Timing Driven Placement", *Proc. ISCAS '95*, pp. 377-380.

[22] M. Sarrafzadeh, D. Knol and G. Tellez, "Unification of Budgeting and Placement", In *Proc. DAC '97*, pp. 758-761.

[23] G. Sigl, K. Doll and F. M. Johannes, "Analytical Placement: A Linear or Quadratic Objective Function?" *Proc. DAC '91*, pp. 57-62.

[24] A. Srinivasan, K. Chaudhary and E. S. Kuh, "RITUAL: A Performance Driven Placement for Small-Cell ICs", *Proc. ICCAD '91*, pp. 48-51.

[25] Takahashi, K.; Nakajima, K.; Terai, M.; Sato, K., "Min-cut Placement With Global Objective Functions For Large Scale Sea-of-gates Arrays.", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, no. 4, April 1995, pp. 434-446.

[26] R. S. Tsay, E. Kuh, "A Unified Approach to Partitioning and Placement", *IEEE Transactions on Circuits and Systems*, Vol.38, No.5, May 1991. pp. 521-633.

[27] R. S. Tsay, E. Kuh, and C. P. Hsu, "Proud: A Sea-Of-Gate Placement Algorithm", *IEEE Design & Test of Computers*,1988, pp 44-56.

[28] Y.-W. Tsay, H.-P. Su, Y.-L. Lin, "An Improved Objective For Cell Placement", *Proc. ASP-DAC '97*, Japan, 1997, pp. 281-284.

[29] J. Vygen, "Algorithms For Large-scale Flat Placement", *Proc. DAC '97*, pp. 746-51.

[30] J. Vygen, *Personal Communication*, November, 1999.

[31] B. X. Weis and D. A. Mlynski, "A new relative placement procedure based on MSST and linear programming. *Proc. ISCAS '87*, pp. 564-567.

[32] H. Xu, "Stochastic penalty function methods for nonsmooth constrained minimization". *Journal of Optimization Theory and Applications*, vol.88, (no.3), Plenum, March 1996. p.709-24.