Global Routing

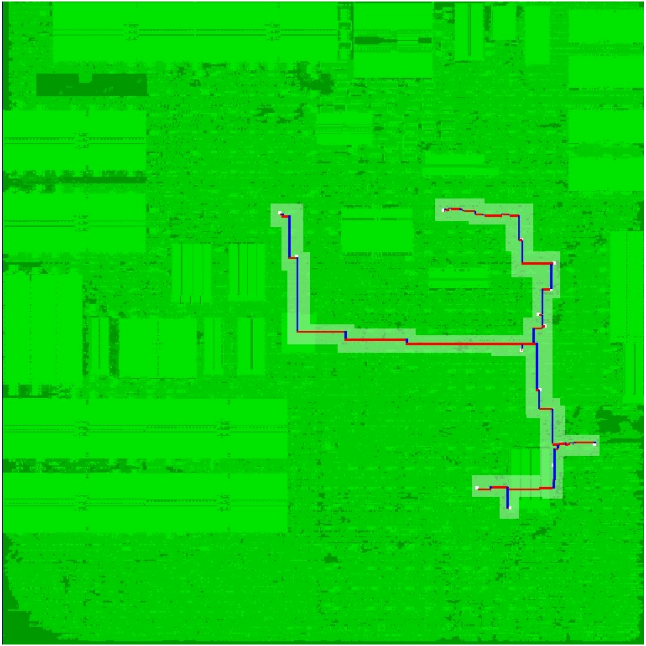
Jens Vygen

Hangzhou, April 2009

Global routing

- contract regions of approx. 100x100 points to a single vertex
- compute capacities of edges between adjacent regions
- pack Steiner trees with respect to these edge capacities
- global optimization of objective functions
- define a detailed routing area for each net according to its Steiner tree

Output of global routing: a corridor for each net



Global routing: simplified problem formulation

Instance:

- a global routing (grid) graph with edge capacities
- a set of nets, each consisting of a set of vertices (terminals)

Task: find a Steiner tree for each net such that

- the edge capacities are respected,
- some objective function (e.g., netlength, yield, or power) is optimized,
- and the timing constraints are met.

Even simple special cases are *NP*-hard!

Fractional relaxation: multicommodity flow problem

Instance:

- an undirected graph G with capacities $u : E(G) \to \mathbb{Z}_+$ and lengths $l : E(G) \to \mathbb{R}$
- a family $\mathcal{N}$ of nets (terminal pairs) with demands $w : \mathcal{N} \to \mathbb{Z}_+$ and weights $c : \mathcal{N} \to \mathbb{Z}_+$

Task: Find a flow f_N for each N of value $w(N)$ such that

$$\sum_{N \in \mathcal{N}} f_N(e) \le u(e) \qquad \text{for } e \in E(G),$$

and

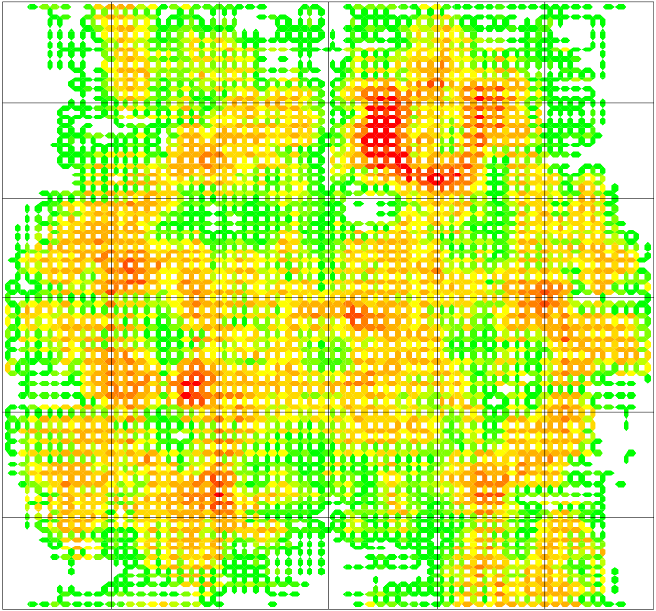
$$\sum_{N \in \mathcal{N}} c_N \sum_{e \in E(G)} l(e) f_N(e) \qquad \text{is minimum.}$$

Multicommodity flows: positive results

- Can be solved by linear programming (but too slow)
- There are combinatorial fully polynomial approximation schemes for the MULTICOMMODITY FLOW PROBLEM: Sharokhi, Matula [1990], Leighton, Makedon, Plotkin, Stein, Tardos, Tragoudas [1991], Plotkin, Shmoys, Tardos [1991], Radzik [1995], Young [1995], Grigoriadis, Khachiyan [1996], Garg, Könemann [1998], Fleischer [2000], Karakostas [2002]
- If edges have sufficient capacity, randomized rounding can be applied to get an integral solution violating capacity constraints only slightly (Raghavan, Thompson [1987,1991], Raghavan [1988])
- This can be applied to Steiner trees instead of paths, works efficiently for large global routing instances (Albrecht [2001])

But this does not take timing constraints and global objectives (power consumption, yield) into account.

Example: global routing congestion map



Constraints and objectives in routing

meet timing constraints

- all signals must arrive in time
- delays depend on electrical capacitances of nets
- capacitance of a net depends on length, width, plane, and distance to neighbour wires (nonlinearly!)

minimize power consumption

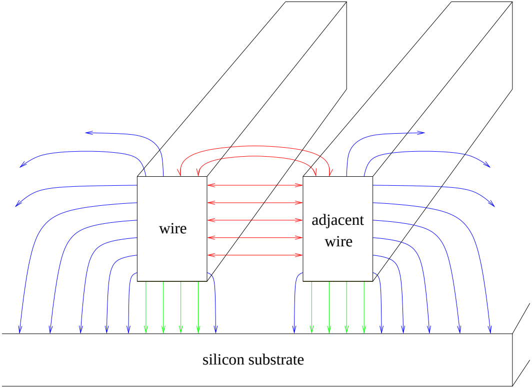
- power consumption roughly proportional to the electrical capacitance, weighted by switching activity

minimize cost

- minimize number of masks (number of routing planes), maximize yield (spreading), minimize design effort

Capacitance estimation

- area capacitance (parallel plate capacitor) – proportional to length times width
- fringing capacitance – proportional to length
- coupling capacitance – proportional to length, inversely proportional to distance to neighbour



Assign extra space to global wires

We assign to each net $c \in \mathcal{C}$ an element of

$$\hat{\mathcal{B}}_c := \Big\{(b, b') \in [0,1]^{E(G)} \times \mathbb{R}_+^{E(G)} : \\ b \text{ incidence vector of a Steiner tree for } c, \\ b_e = 0 \Longrightarrow b'_e = 0 \text{ for all } e \in E(G)\Big\}.$$

- $b_e = 1$ if and only if the Steiner tree for this net uses edge e.
- b'_e is the extra space allocated to $c \in \mathcal{C}$ along edge e.
- Total capacitance of a wire along e can be estimated as a function of b'_e.

Min-max resource sharing

Instance

- finite sets $\mathcal{R}$ of **resources** and $\mathcal{C}$ of **customers**
- for each $c \in \mathcal{C}$:
  - a convex set $\mathcal{B}_c$ of **feasible solutions** (a **block**) and
  - a convex **resource consumption function** $g_c : \mathcal{B}_c \to \mathbb{R}_+^{\mathcal{R}}$
- given by an **oracle function** $f_c : \mathbb{R}_+^{\mathcal{R}} \to \mathcal{B}_c$ with

$$\omega^\top g_c(f_c(\omega)) \leq (1 + \epsilon_0) \inf_{b \in \mathcal{B}_c} \omega^\top g_c(b)$$

  for all $\omega \in \mathbb{R}_+^{\mathcal{R}}$ and some $\epsilon_0 \in \mathbb{R}_+$ (a **block solver**).

Task

- Find a $b_c \in \mathcal{B}_c$ for each $c \in \mathcal{C}$ with minimum **congestion**

$$\max_{r \in \mathcal{R}} \sum_{c \in \mathcal{C}} (g_c(b_c))_r \ .$$

Application to global routing

Given a global routing graph (3D grid with millions of vertices).

- **Customers** = nets (sets of pins; roughly: sets of vertices)
- **Resources** = edge capacities, power consumption, yield loss, timing constraints, ...
- Objective function is transformed into a constraint
- **Block** = (convex hull of) set of Steiner trees for a net, with space consumption for each edge
- Resource consumption is nonlinear (but convex) for yield loss, timing, power consumption
- **Block solver** = approximation algorithm for the Steiner tree problem in the global routing graph (with edge weights)

Yield analysis: critical area

Consider faults caused by particles with size distribution

$$f(r) := \begin{cases} 0, r < r_0 \\ \frac{c}{r^3}, r \geq r_0 \end{cases}$$

for some $r_0 \in \mathbb{R}_+$ smaller than the smallest possible particle that can cause a fault, and c such that $\int_0^\infty f(r)\mathrm{d}r = 1$.

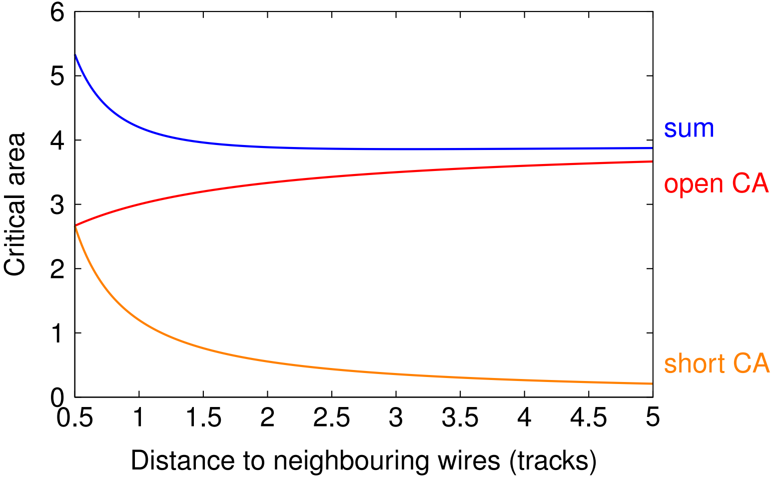
Then the critical area w.r.t. extra material faults on plane z is

$$\mathrm{CA}_{em}^z := \int_x \int_y \int_{t_{em}(x,y,z)}^\infty f(r)\mathrm{d}r\mathrm{d}y\mathrm{d}x,$$

where $t_{em}(x, y, z)$ is the smallest size of a particle that causes an extra material fault at location (x, y, z).

Dependence of critical area on area consumption

Example: Critical area of unit length wire of minimum width



Yield analysis: expected number of faults

Weighted sum of critical areas is used to estimate the number of extra material faults per chip:

$$\mathrm{F}_{em} := \sum_{z} w_{em}^{z} \mathrm{CA}_{em}^{z}$$

Analogously define the number of miss material faults on wire planes, F_{wm}, and on via planes, F_{vm}.

Define the estimated total number of faults per chip as $\mathrm{F} := \mathrm{F}_{em} + \mathrm{F}_{wm} + \mathrm{F}_{vm}$.

The percentage of chips without a fault from one of the above classes is estimated by

$$e^{-\mathrm{F}}.$$

The complement $1 - e^{-\mathrm{F}}$ is called the wiring yield loss.

Modeling yield loss as resource

$$\mathcal{B}_c := \operatorname{conv}(\hat{\mathcal{B}}_c) = \operatorname{conv}\Big(\Big\{(b, b') \in [0,1]^{E(G)} \times \mathbb{R}_+^{E(G)} : \begin{array}{l} b \text{ incidence vector of a Steiner tree for } c, \\ b_e = 0 \Longrightarrow b'_e = 0 \text{ for all } e \in E(G) \end{array}\Big\}\Big).$$

- b'_e is the extra space allocated to net $c \in \mathcal{C}$ along edge e.
- model cost (wiring yield loss) depending on extra space by functions $\gamma_{c,e} : \mathbb{R}_+ \to \mathbb{R}_+$ for $c \in \mathcal{C}$ and $e \in E(G)$.
- Here $\gamma_{c,e}(x)$ is the estimated contribution of edge e, if used by net c with allocated space $\mathrm{minwidth}(c, e) + x$, to the wiring yield loss (similar for power consumption, delay of a path).
- Note that the functions $\gamma_{c,e}$ are convex ($c \in \mathcal{C}$, $e \in E(G)$).
- Resource consumption for new resource γ is given by

$$g_c^{\gamma}(b, b') = \frac{1}{\Gamma} \sum_{e \in E(G): b_e > 0} b_e \cdot \gamma_{c,e}\left(\frac{b'_e}{b_e}\right)$$

for $(b, b') \in \mathcal{B}_c$, where Γ is an upper bound.

Randomized rounding

- Let $\hat{\mathcal{B}}_c \subseteq \mathcal{B}_c$ with $\mathcal{B}_c = \operatorname{conv}(\hat{\mathcal{B}}_c)$.
- Given numbers $x_{c,b} \geq 0$ for all $c \in \mathcal{C}$ and $b \in \hat{\mathcal{B}}_c$ with $\sum_{b \in \hat{\mathcal{B}}_c} x_{c,b} = 1$ for all $c \in \mathcal{C}$.
- Let $\lambda := \max_{r \in \mathcal{R}} \sum_{c \in \mathcal{C}} \sum_{b \in \hat{\mathcal{B}}_c} x_{c,b}(g_c(b))_r$.
- We will compute a solution with

$$\lambda \leq (1+\epsilon) \inf_{b_c \in \mathcal{B}_c (c \in \mathcal{C})} \max_{r \in \mathcal{R}} \sum_{c \in \mathcal{C}} (g_c(b_c))_r$$

  for small $\epsilon > 0$.
- Consider a "randomly rounded" solution, $\hat{b}_c \in \hat{\mathcal{B}}_c$ for $c \in \mathcal{C}$, given as follows.
- Independently for all $c \in \mathcal{C}$ we choose $b \in \hat{\mathcal{B}}_c$ as $\hat{b}_c$ with probability $x_{c,b}$.
- Let $\hat{\lambda} := \max_{r \in \mathcal{R}} \sum_{c \in \mathcal{C}} (g_c(\hat{b}_c))_r$.

Question: can we bound $\frac{\hat{\lambda}}{\lambda}$?

Chernoff bound

Lemma

Let $X_1, \ldots, X_k$ be independent random variables in $[0, 1]$. Let μ be the sum of their expectations, and let $\epsilon > 0$. Then $X_1 + \cdots + X_k > (1 + \epsilon)\mu$ with probability less than $e^{-\mu f(\epsilon)}$, where $f(\epsilon) := (1 + \epsilon)\ln(1 + \epsilon) - \epsilon$.

Note that $f(\epsilon) > 0$ for $\epsilon > 0$.

Proof: Let $\mathrm{Prob}[\cdot]$ denote the probability of an event, and $\mathrm{Exp}[\cdot]$ the expectation of a random variable. Using $(1 + \epsilon)^x \leq 1 + \epsilon x$ for $0 \leq x \leq 1$ and $1 + x \leq e^x$ for $x \geq 0$ we compute

$$\mathrm{Prob}[X_1 + \cdots + X_k > (1 + \epsilon)\mu] = \mathrm{Prob}\left[\frac{\prod_{i=1}^k (1+\epsilon)^{X_i}}{(1+\epsilon)^{(1+\epsilon)\mu}} > 1\right] \leq$$

$$\mathrm{Prob}\left[\frac{\prod_{i=1}^k (1+\epsilon X_i)}{(1+\epsilon)^{(1+\epsilon)\mu}} > 1\right] < \mathrm{Exp}\left[\frac{\prod_{i=1}^k (1+\epsilon X_i)}{(1+\epsilon)^{(1+\epsilon)\mu}}\right] = \frac{\prod_{i=1}^k (1+\epsilon\,\mathrm{Exp}[X_i])}{(1+\epsilon)^{(1+\epsilon)\mu}} \leq$$

$$\frac{\prod_{i=1}^k e^{\epsilon\,\mathrm{Exp}[X_i]}}{(1+\epsilon)^{(1+\epsilon)\mu}} = \frac{e^{\epsilon\mu}}{(1+\epsilon)^{(1+\epsilon)\mu}} = e^{-\mu f(\epsilon)}. \qquad \square$$

(Raghavan, Spencer; see Raghavan [1988] and Chernoff [1952])

Randomized rounding

Theorem

For $r \in \mathcal{R}$ let $\rho_r \geq \frac{(g_c(b))_r}{\lambda}$ for all $b \in \mathcal{B}_c$ and $c \in \mathcal{C}$, and let $\rho := \max_{r \in \mathcal{R}} \rho_r$. Let $\Omega := \rho \max \left\{ 1, \ln \left(\sum_{r \in \mathcal{R}} e^{1 - \frac{\rho}{\rho_r}} \right) \right\}$ and $\delta := (\Omega + e - 2)\sqrt{\frac{\Omega}{f(\Omega + e - 2)}}$.
Then $\hat{\lambda} \leq \lambda(1 + \delta)$ with positive probability.

Proof (sketch):

For each resource $r \in \mathcal{R}$, apply the above Chernoff bound to the independent random variables $\frac{(g_c(\hat{b}_c))_r}{\rho_r \lambda}$, $c \in \mathcal{C}$. $\square$

(Müller, V. [2008]; see Raghavan [1988])

In practice:

Some violations occur, are fixed by “rip-up and re-route”

Critical area after detailed routing

Chip	Tech.	#Nets	Netl. Opt.	Yield Opt.	
Edgar	Cu08	772,000	0.10493	0.08586	(–18.2%)
Hannelore	Cu08	140,000	0.01543	0.01027	(–33.4%)
Paul	Cu08	68,000	0.00568	0.00402	(–29.2%)
Monika	Cu11	1,502,000	0.09505	0.08055	(–15.3%)
Garry	Cu11	827,000	0.08017	0.06714	(–16.3%)
Heidi	Cu11	777,000	0.05804	0.04965	(–14.5%)
Elena	Cu11	421,000	0.03314	0.02966	(–10.5%)
Lotti	Cu11	132,000	0.00688	0.00575	(–16.4%)
Ingo	Cu11	58,000	0.00505	0.00392	(–22.4%)
Bill	Cu11	11,000	0.00833	0.00376	(–54.9%)
Total			**0.50190**	**0.41419**	**(–17.5%)**

(Müller [2006])

Summary

- Global routing is a generalization of integer multi-commodity flow
- fractional solutions are useful, can be made integral by randomized rounding (with some loss)
- linear programming too slow
- combinatorial fully polynomial time approximation schemes much better
- multi-terminal nets, nonlinear constraints and objectives (like yield, power consumption, timing) can be modeled in terms of the min-max resource sharing problem
- tomorrow: an efficient algorithm for this problem