VFGR: A Very Fast Parallel Global Router with Accurate Congestion Modeling

Zhongdong Qi, Yici Cai, Qiang Zhou, Zhuoyuan Li and Mike Chen

1 Tsinghua University, Beijing, China
2 Nimbus Automation Technologies, Shanghai, China
Outline

- Background and Motivation
- Proposed Congestion Model
- Parallel Global Routing
- Experimental Results
- Conclusions
Global Routing in Design Flow

- Routing: complex and important
  - Determines geometry and location of interconnect features under several constr.
  - Largely affects performance, power and yield.
- Global routing (GR) plans tree topologies.
- Detailed routing (DR) constructs wires and vias.
GR Formulation and Research Status

- **GR Formulation**
  - Input: routing graph, a set of nets
  - Output: routing trees for nets
  - Objectives: congestion, wirelength, via count, etc.

- Long research history, great progress recently
  - High performance and quality routers
    - FGR, BoxRouter 2.0, NTHU-Route 2.0, GRIP, NCTU-GR 2.0 …
Challenges for Global Routing

- Technology nodes get smaller
  - More metal layers, e.g. 6(90nm) → 9(65nm) → 12(45nm) ...
  - Varying metal widths
    - Fat vias, more stacked vias
  - More design rules
    - More resource consumption by global and local connections

- Larger design size and problem complexity
  - Increased chip dimension and nets, 3-D problem

![Fat via](image1.png)

![65nm layer stack](image2.png)
Facing the Challenges

- An practical congestion model
  - Captures the local congestion by vias & local connections
  - Explicitly models most influential design rules
- A multi-threaded global routing algorithm
  - A global routing framework easier to be parallelized
  - Region level parallelism and net level parallelism
Outline

- Motivation and Background
- Proposed Congestion Model
- Parallel Global Routing
- Experimental Results
- Conclusion
Real Congestion in Sub-65nm Technologies

- Not measured in conventional congestion model
- Make a big gap between global routing and detailed routing
Proposed Concept: Pass-through Capacity and Demand

Use pass-through capacity/demand to model intra-gcell congestion
Proposed Model: Pass-through Capacity and Demand (cont’d)

- **Capacity**: available tracks and partial tracks
- **Demand contributors**
  - Fat via enclosure and stacked via enclosure
    - Affected by MinArea, EOL-Spacing and normal spacing
  - Local net connection
    - Net connection tree: RSMT generated by FLUTE
    - Affected by MinArea and EOL-Spacing
  - Global net segments

<table>
<thead>
<tr>
<th>Feature</th>
<th>Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat via enclosure</td>
<td>3</td>
</tr>
<tr>
<td>Stacked via enclosure or local net connection</td>
<td>(minArea/width+2*eolSpace)/gcellWidth (if necessary)</td>
</tr>
<tr>
<td>Segment crossing gcell</td>
<td>1</td>
</tr>
<tr>
<td>Segments connecting to gcell</td>
<td>max ( {N_i, N_r} )</td>
</tr>
</tbody>
</table>
Proposed Congestion Model in 3-D Routing Graph

Compatible with widely used path search algorithms in GR, e.g. pattern routing, maze routing, layer assignment etc.
Outline

- Motivation and Background
- Proposed Congestion Model
- Parallel Global Routing
- Experimental Results
- Conclusion
Global Routing Framework

- Take negotiated-congestion routing as foundation.
  - Adopted most in state-of-the-art global routers.

- Observation
  - Smaller nets: in local region and lower layers
  - Larger nets: in larger scope and higher layers
  - Smaller nets has less flexibility, larger ones more

- Hierarchical global routing framework
  - From local region nets (lower level) to global region nets (higher level)
  - Progressively construct the routing solution using negotiated-congestion routing
Global Routing Framework (cont.)

- Progressively construct the routing solution
  - Multiple hierarchies with different-size regions
  - From bottom level to top level
  - In each level, all the nets inside regions are routed using negotiated-congestion routing
Issue to Handle

- Region restriction
  - No enough resources in some regions
  - Congestion and(or) detours
  - Solution: deferring congested nets and detoured nets to next level

Congestion due to region restriction
Detour due to region restriction
Global Routing Flow

1. **Net decomposition and hierarchy construction**
   - $i = 0$

2. **Level-i routing with region-level parallelism**

3. **Congested nets and detoured nets deferring**
   - **Top level?**
     - **Y**
       - **Top level routing with net-level parallelism**
       - **End**
     - **N**
       - $i = i + 1$

The flowchart illustrates the process of global routing, starting with net decomposition and hierarchy construction, followed by progressive levels of routing until the top level is reached or all nets are resolved.
Region Level Parallelism

- In lower levels, routing nets in different regions are independent.
- Routing in each region is constructed as a task.
Net Level Parallelism

- In top level, routing each net is regarded as a task.
- Dynamically select nets; bounded box A* search;
- Compatibility: path search bbox overlap-free.
Experimental Setup

- **Benchmarks**: DAC 2012 Benchmark Suite
  - 1X for M1-M4, 2X for M5-M7, 4X for M8-M9
  - Added 65nm design rules

- **Machine**: Intel 8-core 2.40GHz CPU & 24GB memory

<table>
<thead>
<tr>
<th></th>
<th>#Net</th>
<th>#Layer</th>
<th>Grid</th>
<th>#G-Net</th>
<th>Grid</th>
<th>#G-Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>superblue2</td>
<td>990899</td>
<td>9</td>
<td>770 x 891</td>
<td>698751</td>
<td>2738 x 3960</td>
<td>910675</td>
</tr>
<tr>
<td>superblue3</td>
<td>898001</td>
<td>9</td>
<td>800 x 415</td>
<td>554046</td>
<td>2845 x 1845</td>
<td>796005</td>
</tr>
<tr>
<td>superblue6</td>
<td>1006629</td>
<td>9</td>
<td>649 x 495</td>
<td>643532</td>
<td>2308 x 2200</td>
<td>912526</td>
</tr>
<tr>
<td>superblue7</td>
<td>1340418</td>
<td>9</td>
<td>499 x 713</td>
<td>899583</td>
<td>1775 x 3169</td>
<td>1236238</td>
</tr>
<tr>
<td>superblue9</td>
<td>833808</td>
<td>9</td>
<td>625 x 570</td>
<td>528664</td>
<td>1515 x 2534</td>
<td>771628</td>
</tr>
<tr>
<td>superblue11</td>
<td>935731</td>
<td>9</td>
<td>631 x 878</td>
<td>678379</td>
<td>2244 x 3903</td>
<td>879756</td>
</tr>
<tr>
<td>superblue12</td>
<td>1293436</td>
<td>9</td>
<td>444 x 518</td>
<td>919016</td>
<td>1579 x 2303</td>
<td>1213706</td>
</tr>
<tr>
<td>superblue14</td>
<td>619815</td>
<td>9</td>
<td>406 x 473</td>
<td>433978</td>
<td>1444 x 2103</td>
<td>579239</td>
</tr>
<tr>
<td>superblue19</td>
<td>511685</td>
<td>9</td>
<td>321 x 518</td>
<td>339259</td>
<td>1142 x 2303</td>
<td>470070</td>
</tr>
</tbody>
</table>
Experimental Results (1)

- GR performance and solution quality
  - Benchmark: DAC 2012 benchmark suite
  - Compared with: NCTU-GR 2.0 and BFG-R

Three routers all eliminate overflow for all testcases.
Comparable or better wirelength and via count
About 6X speed up for parallelization

* NCTU-GR 2.0 and BFG-R are not multi-threaded.
Experimental Results (2)

- Effectiveness of proposed congestion model
  - GR by BFG-R / VFGR + DR by a commercial drouter
  - Benchmark: fine-gcell DAC 2012 benchmark suite
- Global routing results:

Performance on designs with large routing grid:
Parallelized router is 8 times faster than BFG-R
Experimental Results (3)

- Effectiveness of proposed congestion model
  - GR by BFG-R / VFGR + DR by a commercial drouter
  - Benchmark: fine-gcell DAC 2012 benchmark suite
  - Detailed routing results:
    - 59% fewer design rule violations
    - 6% shorter DR wirelength
    - 9% fewer DR via count
    - 51% shorter DR runtime

Captures DR congestion
Guides detailed router better
Conclusion

- Proposed pass-through capacity and demand to model intra-gcell congestion, better correlated to DR resource consumption.
- Considering DR effects in GR leads to much shorter DR runtime and better DR results.
- Hierarchical global routing framework, which enables easier parallelization.
- Achieved comparable GR solution quality with NCTU-GR 2.0 and BFG-R, and near 6X speedup for parallelization.
Thank you!
Net Decomposition

- Use both RSMT and RMST
  - MST edges as sub-nets; small cost for Steiner nodes.
  - Flexibility of path search; short wirelength
  - Refer sub-net as “net” in the following pages.
Nets in different levels

- Order of different levels
  - Lower level nets are routed ahead of higher level nets
  - Higher level nets may have less flexibility
  - Solution: all the nets inside the current region can be rerouted using negotiated-congestion routing

![Diagram showing nets in different levels](image-url)