Effective Resource Model and Cost Scheme for Maze Routing in 3D Global Routing

Zirui Li[1], Jianwang Zhai[1], Zixuan Li[1], Zhongdong Qi[2], Kang Zhao[1]
[1]Beijing University of Posts and Telecommunications [2]Xidian University
{lzr_offical, zhaijw, gxydbc, zhaokang}@bupt.edu.cn, zdqi@xidian.edu.cn

Introduction

Routing is an essential step to VLSI design closure whose flow is generally divided into global routing and detailed routing. Global routing is to find rough routes within the routing area under a series of constraints, aiming at the minimum cost while taking into account detailed routability. However, the initial routing might not meet all the requirements due to limited routing resources. In such cases, the global router needs to remove some existing connections and then re-find and re-establish better routes to meet the requirements, which is known as rip-up and reroute (RRR). As shown in Fig. 1, the runtime percentage of RRR in global routing is very large, and more efficient models and algorithms are urgently needed.

Maze routing [1] is one of the commonly operated methods in the RRR iterations. This method runs on a grid map representing the routing region and amount of routing resources, aiming to find the shortest path under routing resource constraints. MGR [2] adopts a multi-level framework and two-stage 3D maze routing method to respectively plan and search for routes at coarse and fine granularity. It greatly balances the performance and runtime. CUGR [3] proposes a multi-level 3D maze routing algorithm with a probability-based cost scheme that better considers detailed routability, and applies it to the RRR iterations.

Existing 3D global routers optimize algorithms and frameworks and have achieved remarkable results. However there is room for improvement in accurate modeling of routing resources and cost function in multi-level 3D maze routing. Taking account of distribution while evaluating the resources or more precise value setting of parameters may bring performance in runtime and solution quality.

In this work, we propose an effective resource model and congestion adjusting method, to pursue higher efficiency and better results.

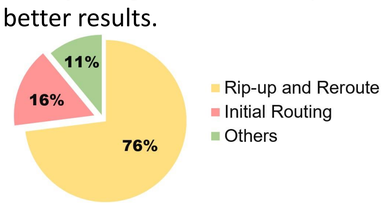

Fig.1 Breakdown runtime results on ICCAD2019 benchmarks [4] by CUGR [3] given in [5].

Optimization of Multi-level 3D Maze Routing

In original Multi-level 3D maze routing, CUGR directly compute the resource of a coarsened cell by the arithmetic mean value of resources of all the G-cells within:

$$R(A)=\frac{\sum_{i=0}^{s-1}\sum_{j=0}^{s-1} r\left(v_{ij}\right)}{s^2}.$$

G-cells aggregate into coarsened cells and coarsened edges are formed by the interconnections between coarsened cells. The closer the G-cells within coarsened cells are to the connection, the more important its resources are. If the intersecting boundary resources of two adjacent coarsened cells are insufficient to form a continuous edge, regardless of the abundance of resources in their respective intermediate regions, they cannot establish a path. Therefore, the G-cell resource in each coarsened cell should be given corresponding weights according to its location.

A wire edge is split into several pairs of adjacent coarsening units, higher weights should be assigned to resources of G-cells near the connection. Assuming A and B is a pair of horizontally adjacent coarsened cells, the weight $w(u_{ij})$ of resources in G-cell u_{ij} can be computed as:

$$w(u_{ij})=\begin{cases}\frac{x_{ij}-x_{low}}{s_x}, u_{ij}\in A\\ \frac{x_{high}-x_{ij}}{s_x}, u_{ij}\in B\end{cases},$$

$$w(u_{ij})=\begin{cases}\frac{y_{ij}-y_{low}}{s_y}, u_{ij}\in A\\ \frac{y_{high}-y_{ij}}{s_y}, u_{ij}\in B\end{cases}.$$

In layers routed horizontally, x_{high} and x_{low} represent the largest and smallest value of horizon abscissa in a pair of coarsened cells, and s_x refers to the coarsening scale. For s G-cells, the sum of weight changes from $s\times1$ to $\frac{1}{s}+\frac{2}{s}+\ldots+\frac{s-1}{s}+1=\frac{s+1}{2}$. It is necessary to multiply by $\frac{2s}{s+1}$ to calculate the total resources of coarsened cells in wire edges:

$$R(A)=\sum_{i=0}^{s-1}\sum_{j=0}^{s-1} r(u_{ij})\times w(u_{ij})\times\frac{2s}{s+1}.$$

For vias, higher weight should be assigned to the resources near the edge area. Suppose that A and B is a pair of coarsened cells in a via, respectively located in layers routed horizontally and vertically. Using (x_{mid}, y_{mid}) to represent the center coordinates of a coarsened cell, the weight $w(u_{ij})$ of resources in G-cell u_{ij} is calculated using:

$$w(u_{ij})=\begin{cases}\frac{|x_{ij}-x_{mid}|}{s_x}, u_{ij}\in A\\ \frac{|y_{ij}-y_{mid}|}{s_y}, u_{ij}\in B\end{cases}.$$

Similarly, the weight sum of s G-cells in a coarsened via changes from $s\times1$ to $\frac{\frac{s+1}{2}}{s}+\ldots+\frac{2}{s}+\frac{1}{s}+\frac{2}{s}+\ldots+\frac{\frac{s+1}{2}}{s}=\frac{\frac{s}{2}+1}{2}$. The resources of coarsened cells in via edges can be computed as:

$$R(A)=\sum_{i=0}^{s-1}\sum_{j=0}^{s-1} r(u_{ij})\times w(u_{ij})\times\frac{2s}{\frac{s}{2}+1}.$$

Fig. 3 and Fig. 4 give the example resource weight distribution of a pair of coarsened cells in a wire and via.

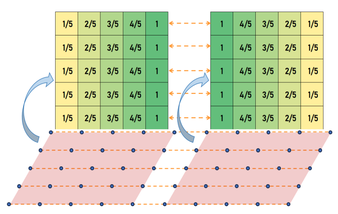

Fig.3 Example of weight distribution in coarsened edges.

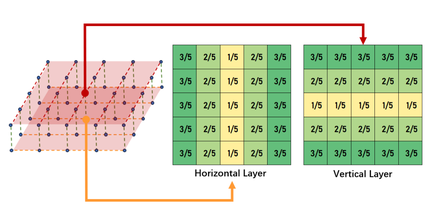

Fig.4 Example of weight distribution in coarsened vias.

Preliminaries

Global Routing

Global routing works on a set of horizontally and vertically evenly distributed G-cells. By considering each G-cell as a vertex $v\in V$ and connections between adjacent G-cells as an edge e $\in E$, the routing area can be formed as a grid graph $G(V,E)$. The connections between G-cells on the same layer and different layers are called wire edges and via edges, respectively. The capacity of edges is defined as the number of tracks going through. The number of tracks that remain and are occupied are called the edge's resource and demand. Finding a route connecting all the pins under design rules with minimum cost while meeting the capacity requirements is the main purpose of global routing.

Multi-level 3D Maze Routing

Multi-level 3D maze routing introduces a technology called coarsening to narrow the 3D search space and divide RRR into two levels in different granularity. Fig. 2 shows the procedure of multi-level 3D maze routing. First, blocks of multiple G-cells are compressed at the scale of s into coarsened cells which reconstruct the original grid graph into a coarsened grid graph. The resource of a coarsened cell is computed as the average resource of all the G-cells within it.

Then, coarse-grained maze route planning is performed on the coarsened grid graph to find the rough search area of the optimal route. In this way, the fine-grained maze router will find the minimum cost route within the planned area instead of directly searching the whole 3D graph.

Cost Scheme

The proposed methodology is based on the competitive global router, CUGR [3], whose primary notations are listed in TABLE I. CUGR not only takes into account the congestion situation during the routing process, but also dynamically adjusts the sensitivity to congestion in different RRR iterations. Take the wire edge as an example, the cost function is:

$$cost_{w(u,v)}=wl(u,v)+eo(u,v)\times lg(u,v),$$

$$eo(u,v)=wl(u,v)\times\frac{d(u,v)}{c(u,v)}\times uoc,$$

$$lg(u,v)=\frac{1}{1+e^{(slope\times r(u,v))}}.$$

The parameter $slope$ determines to what extent of resource scarcity lead to the rapid increase of the congestion cost. In other words, $slope$ determines the router's congestion sensitivity. Defining the possibility of overflow as $1-\frac{r(u,v)}{c(u,v)}$, the congestion weight $wcg_w(u,v)$ can be expressed as:

$$wcg_w(u,v)=uoc\times\left[1-\frac{r(u,v)}{c(u,v)}\right]\times lg(u,v).$$

TABLE I Notations

Symbol	Definition
u,v	G-cells
(u,v)	A wire edge or via edge with u and v as endpoints
$c(u,v)$	Capacity of edge (u,v)
$d(u,v)$	Demand of edge (u,v)
$r(u,v)$	Resource of edge (u,v)
$wl(u,v)$	Wire length of a wire edge (u,v)
$eo(u,v)$	Expectation of overflow of a wire edge (u,v)
$lg(u,v)$	Logistic coefficient of an edge (u,v)
uoc	Unit overflow cost
$slope$	An adjustable parameter of overflow sensitivity
s	The scale of coarsening

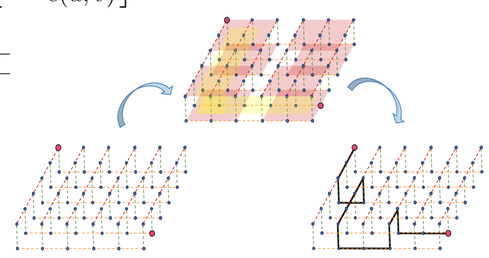

Fig.2 The illustration of multi-level 3D maze routing.

Dynamic Adjustment Method to Congestion Sensitivity

We propose a more accurate parameter update method, which is used to adjust the sensitivity to congestion of different RRR iterations. To ensure that the congestion cost starts increasingly weighted from the same level with the reduction of resources in each RRR iteration, $slope$ and resource $r(u,v)$ need to satisfy: $slope\times r(u,v)=\lambda$.

λ is a given constant that determines the level at which the congestion cost weights begin to increase. The parameter il is the number of RRR limits, and i is the number of RRR iterations. For the i^{th} iteration, the weight of congestion cost $wcg_w(u,v)$ should start increase when the resource $r(u,v)=i_l-i$. To unify the starting points in each iteration that $wcg_w(u,v)$ increasing, $slope$ is calculated as: $slope=\lambda/(i_l-i)$.

When $i_l=4$, the logistic function $lg(u,v)$'s curves in each iteration are shown in Fig. 5. Compared with CUGR, our first RRR iteration reduces the sensitivity to a certain extent.

Therefore, lower-cost path will be chosen in the first RRR iteration. It also can be seen that the sensitivity to congestion in the last two RRR iterations is almost unchanged. Consequently, the congestion degree of the routing result will not change much compared to the original version.

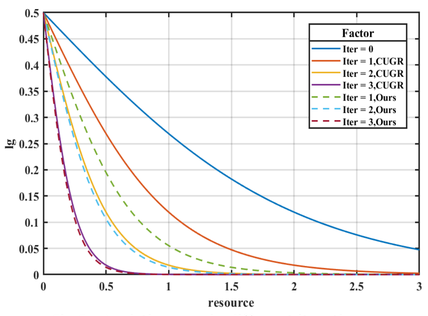

Fig.5 Logistic curve in different iterations.

Evaluation

Our methods is implemented in C++ based on CUGR [3]. Experiments are conducted on a Linux platform with Intel i9-13905H @5.4GHz CPU and 32GB memory. ICCAD2019 benchmarks [4] is adopted to estimate the performance and compare the routing efficiency and quality with CUGR. A weighted sum score s is used to evaluate the quality of the global routing results, which is defined as:

$$s=\alpha W+\beta V+\gamma S.$$

W means the wirelength, V is the number of vias, and S is the number of short violations, α, β, and γ are weighting factors set to 0.5, 4, and 500, according to [4]. To measure the improvement in efficiency more fairly and accurately, each case is run 5 times and averaged in the same environment.

As shown in TABLE II, our methodologies achieves a 6.64% speed-up in the RRR stage to CUGR, and the overall runtime of global routing is reduced by 5.69%. This is mainly due to our improvement to multi-level 3D maze routing. Since the spatial distribution of resources is fully considered in the RRR stage, the elimination of suboptimal options performs faster. Consequently, the solution space is optimized to a certain extent which leads to less runtime.

TABLE II also gives the comparison for the global routing results of all benchmarks. Overall, the wirelength is shortened by 0.04%, and the number of vias is reduced by 0.27%, which is mainly due to our optimization of the cost scheme. The maze router is given more routing options with lower costs because the weight of congestion cost in the first RRR iteration is decreased. In the remaining iterations, the sensitivity of congestion is close to CUGR, which ensures minimal short violations due to congestion. As a result, the overall short violations are reduced by 2.29% and the average score is reduced by 0.07%.

TABLE II Comparison of Decomposed GR Scores and Runtime

Design	Wirelength		# Vias		# Shorts		Score		RRR Runtime (s)		Overall Runtime (s)	
	CUGR	Ours	CUGR	Ours	CUGR	Ours	CUGR	Ours	CUGR	Ours	CUGR	Ours
18test5	27,007,300	27,009,300	855,453	854,563	0	0	16,925,500	16,922,900	21.143	18.660	34.943	31.457
18test5m	27,871,000	27,856,900	803,626	798,200	3,141	3,204	18,720,500	18,723,300	38.966	35.270	44.368	40.922
18test8	64,347,700	64,352,800	2,175,330	2,173,670	0	0	40,875,200	40,871,000	127.779	122.347	151.326	146.208
18test8m	64,700,300	64,574,100	1,955,550	1,950,320	5,527	5,738	42,935,600	42,957,100	151.379	139.463	168.172	156.627
18test10	66,772,100	66,748,300	2,309,520	2,306,100	0	0	42,624,100	42,598,500	143.817	139.102	167.954	163.016
18test10m	70,863,900	70,752,300	20,780,700	2,070,720	652	566	44,070,200	43,942,000	206.627	194.642	223.315	210.950
19test7	118,717,000	118,724,000	3,124,710	3,120,790	0	0	71,857,300	71,845,100	344.578	325.069	392.081	373.959
19test7m	106,995,000	107,054,000	30,713,500	30,560,300	4,367	4,284	67,966,200	67,893,100	209.886	203.209	249.219	236.754
19test8	181,899,000	181,883,000	57,481,700	57,452,900	0	0	113,942,000	113,922,000	178.908	160.141	249.465	230.910
19test8m	177,693,000	177,569,000	55,993,700	55,786,900	6,479	6,176	114,484,000	114,187,000	444.648	417.303	503.840	477.327
19tes9	181,899,000	181,883,000	57,481,700	57,452,900	0	0	175,532,000	175,496,000	241.538	220.801	365.717	343.861
19test9m	267,648,000	267,471,000	9,342,080	9,296,560	3,741	3,749	173,062,000	172,796,000	408.035	395.224	533.633	514.578
Average	**-0.04%**		**-0.27%**		**-2.27%**		**-0.07%**		**-6.64%**		**-5.69%**	

Conclusion

In this work, we propose an effective resource model for the RRR iterations of global routing. First, the multi-level 3D maze routing is optimized by weighing resources of G-cells according to their distributions. Then, we redesign the parameter's updating method in the cost scheme to better adjust the sensitivity to congestion. As a result, our method achieves both routing efficiency and quality improvements on competitive global routers.

Reference

[1] J. Soukup, "Fast Maze Router," in ACM/IEEE Design Automation Conference (DAC). IEEE Computer Society, 1978, pp. 100–101.
[2] Y. Xu and C. Chu, "MGR: Multi-level Global Router," in IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2011, pp. 250–255.
[3] J. Liu, C.-W. Pui, F. Wang, and E. F. Y. Young, "CUGR: DetailedRoutability-Driven 3D Global Routing with Probabilistic Resource Modelx," in ACM/IEEE Design Automation Conference (DAC), 2020, pp. 1–6.
[4] D. Sergei, V. Alexander, W. Lutong, and X. Bangqi, "2019 CAD Contest: LEF/DEF Based Global Routing," Jan 2019.
[5] J. Liu and E. F. Young, "EDGE: Efficient DAG-based Global Routing Engine," in ACM/IEEE Design Automation Conference (DAC), 2023, pp. 1–6.