

Early Congestion Detection and Mitigation in a Macro-Rich 28nm ASIC Design

Gundelli Srikanth¹, Harsha Vardhan Reddy. G¹, Srinidhi Prabhakar²
Vijay Vallabhuni^{2,*}

¹*Silica Launch, Bangalore, India,*

²*Astrasilica, Bangalore, India,*

¹*gundellisrikanthyadav@gmail.com, ¹harshareddygajula@gmail.com,*

²*srinidhipyadav@gmail.com, ^{2,*}vvijay@astrasilica.com*

(Received 12 August 2025; Revised 04 November 2025; Accepted 21 December 2025;
Published 02 January 2026)

Abstract

One of the most important issues of current ASIC physical design, especially the macro-dominated layout, is routing congestion. Congestion related problems are known to be detected at the end of the routing process of many industrial design flows, leading to numerous design cycles, longer turn around time, and timing closure variability. The paper describes a preliminary congestion prediction and mitigation algorithm to a macro-rich 28-nm ASIC architecture. The suggested solution is based on the global routing congestion measures, macro flyline, pin density assessment, and boundary logic allocation to detect the congestion-prone areas before detailed routing. Using the early indicators of congestion to correlate with downstream routing failures and setup timing violations, high-risk regions are identified with high confidence. Specific corrective measures such as refinement of macro placement, redistribution of boundary cells.

Keywords: Early Congestion Detection, Macro Flyline and Pin Density Analysis, Placement-Stage Congestion Analysis, Routability and Timing Closure, Routing Congestion Mitigation

1 Introduction

One of the challenges of advanced physical design is routing congestion that is a constraint to routability and timing closure as technology scale increases. This is especially severe in the 28-nm node, in designs with lots of macros, where macros hardness is a constraint on routing resources and routing demand is concentrated at small locations. The traditional flows are typically employed to solve congestion during or after routing, but remedies that are provided at a late stage cannot

successfully solve any congestion caused by early-stage placement decisions, e.g. macro placement and interface clustering. In these cases, routing detours lead to an increase in routing wirelength and routing parasitics that lead to timing closure instability.

1.1 Significance of Congestion Analysis

Due to the increasing scale of technology, interconnect has emerged as the most prominent performance, power and reliability factor in modern VLSI designs. Logic density is increasing at an increasing rate, however, routing resources are not increasing at the same rate, causing severe issues of congestion, particularly at macro-rich designs. At the 28-nm technology node, we no longer have a local routing problem but a design disadvantage that directly impacts timing closure, signal integrity and design convergence in general.

The conventional physical design approaches typically viewed congestion as a downstream problem, and attempted to remove routing overflows at global or detailed routing stages. Such corrective actions are possible late in the design phase, which may result in large engineering change orders (ECOs), routing perturbations, and multiple routing runs, which causes a huge increase in the overall design turnaround time and reduction in convergence predictability (Lin et al. 2020; Pan and Chen 2021).

1.2 Limitations of Post-Route Congestion Fixes

The congestion problem is generally addressed in more detail during routing or after placement in more detailed physical design flows, when routing overflows and design rule violations have been identified. Some of the corrective measures, which may involve the incremental cell spreading, repositioning of metal layers or a limited macro movement is normally adopted at this stage. However, late-stage alterations may result in time-dependent routes and add parasitic influence, and present new establishment violations and signal integrity problems. This renders the resolution of congestions a reactive and iterative process, reducing predictability of designs as well as convergence in particular macro-dominated design.

1.3 Motivation for Early Congestion Detection

The purpose of early congestion detection is to detect and mitigate routing-critical areas in the early stages of placement, where the design has the most flexibility. Early-stage congestion measurement methods such as probabilistic global routing abstractions, Rents-rule-directed complexity analysis and tile-based demand-supply modeling have demonstrated that the congestion during placement is a good predictor of the ultimate routing bottlenecks. These techniques allow routing critical areas to be identified long before detailed routing, when there is still plenty of flexibility in the design (Lin et al. 2020; Yu et al. 2022; Wang et al. 2021; Xu and Kahng 2022). Using early congestion signals-macro flylines, pin density and boundary-cell

clustering designers can be proactive in their placement decisions to avoid extreme routing failures. Such a change in the reactive to proactive congestion control is crucial to attaining predictable convergence within the advanced-node designs.

1.4 Scope and Contributions of the Proposed Study

This paper makes the following key contributions: Using the 28-nm Engineering Change Design Methodology (ECDM) block as an archetype of an industrial example, this paper demonstrates that the signature of congestion at the early stages of placement is closely correlated with post routing downstream routing failure, including persistent setups, excessive routing detours and design rule breakages . It presents a viable, ICC2-compatible algorithm to early analysis of congestion which is a combination of macro flyline analysis, distribution based on boundary cells, and density sensitive placement analysis. The first one is that it validates targeted mitigation strategies that are performed early in routing without overflow and enhance timing stability without disruptive late-design changes

2 Design Methodology

2.1 Block Overview: ECDM

The Engineering Change Design Methodology (ECDM) block is a dense subsystem that is timing-critical and built in 28-nm CMOS and is optimized to have timing closure through repeaters. It has approximately 75,000 leaf cells in an area of $313,567 \mu\text{m}^2$ core and is highly macro-dominated. The macros of six hard memories occupies around 250, 205 205 mm^2 with nearly 80 per cent of the area occupied by the macro core as shown in Fig. 1.

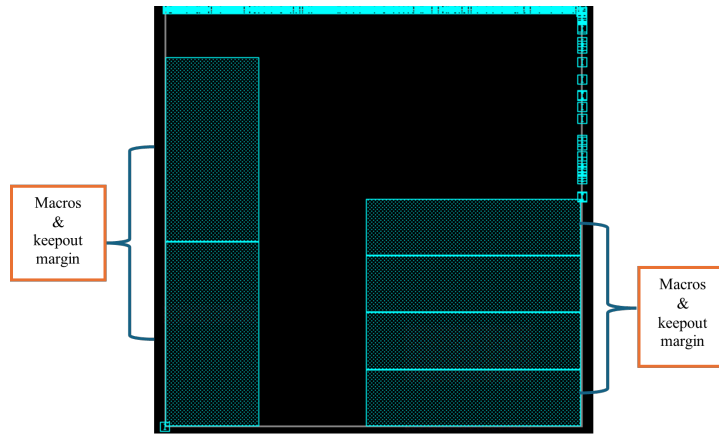


Fig. 1. Floorplan of ECDM showing abutted macro cluster, halo regions, and boundary cell placement.

The two highly sensitive global nets of the block, both with fanouts greater

than 1,000, have a target frequency of 1.2 GHz, and presented a severe timing and routing challenge. The positioning of the high-drive repeaters is very minimal with a 20 percent effective whitespace only. Consequently, RTL to GDSII implementation is anchored on careful macro-conscious keeping out and location control to avoid routing congestion in the fine channels between the dominating macros.

2.2 Macro and Boundary Cell Organization

The Engineering Change Design Methodology (ECDM) macros are placed in a location which is in accordance to the best practices of macro placement like proper alignment, spacing and orientation necessities. In addition, one has the flyline analysis which is used to direct the macro positioning in such a way that long routing crossing are minimized and unnecessary high connectivity paths are minimized. The halo areas are used to contain all the macros to prevent intrusion of standard cells besides ensuring macro routability. Even though halos are used to protect macro interfaces, they also reduce the space available to place standard cells bringing standard-cell density nearer to macro edges. The macros are surrounded by boundary cells, which are laid to bridge macro signals to logic subsequently and the density of the pins on the macro boundary of the abutted macro cluster is high, particularly the pins on the macro cluster being abutted.

2.3 Floorplan Characteristics

The ECDM floorplan is macro dominated and routing resources are mostly determined by macro placement as opposed to even cell distribution. This abutted macro cluster is like a routing barrier, which makes interconnections traverse narrow routing channels along macro edges. Standard cells are restricted to non-halo regions and therefore the high cell density is localized around the macros. This uneven distribution generates routing stress areas at the very start of the placement phase, before the finer routing has started.

2.4 Routing Constraints

The problem with routing in ECDM is primarily the interaction between abutted macros, halo regions, and high density of pin density on high-density boundaries of the macro. These limitations decrease the routing tracks available and limit the routing escape routes around macros. Besides that, the routing flexibility on lower metal layers is further reduced by the power grid structures around macro boundaries. This makes ECDM congestion very sensitive to early placement decisions and it is important that early congestion analysis be taken to avoid downstream routing and timing problems.

3 Sources of Congestion

3.1 Macro Flyline-Induced Congestion

Macro flylines are commonly utilized as a simple precursor to the connectivity strength and routing demand in placement phases (Lin et al. 2020; Yu et al. 2022). In previous work, it has been demonstrated that localized routing congestion typically occurs when some regions are concentrated (Kahng et al. 2021; Pan and Chen 2021). The flyline analysis in ECDM shows that many of the connections are pulled to certain particular macro interfaces, especially in the abutted macro cluster. This convergence pre-emptively routing demand in the future into limited areas around macro edges forming hotspots of congestion prior to the world routing.

3.2 Boundary Cell Clustering

Redundant routing congestion is often due to overuse of routing resources by local connections leaving no capacity to support the rest of the nets (Kahng et al. 2021; Lin et al. 2020). The effects are further enhanced in areas where the interface is complex (Yu et al. 2022). Boundary cells in ECDM that are put around abutted macros cluster on common macro edges. This clustering makes the pin density of the locality higher and results in routing resource starvation at the macro boundaries.

3.3 Pin Density Impact

The pin density is known to be a cause of routing congestion especially in designs that are rich in macros whereby routing resources around the macro boundaries are sparse (Kahng et al. 2021; Yu et al. 2022). According to previous research, congestion becomes worse when pin concentration between the macros and standard cells is more than the routing capacity of the locality which can lead to directional routing stress in many cases (Yu et al. 2022; Wang et al. 2021). At the ECDM block, the pin density is donated by the macro interface pins, boundary cells and the pins that are placed side by side with the macro halos. Halo constrained standard cells will be concentrated along macro boundaries and will create local pin concentration. This density of macros and standard cells in combination forms many short and competing connections that have to escape into few routing channels and as a result, early congestion is created on the edges of macro cells.

3.4 Power Grid Interaction

Power distribution network (PDN) is a spread of VDD and VSS metal stripes forming a dense mesh that covers the entire block to provide the current. At high-technology nodes like 28 nm, the PDN takes up a large part of routing resources and competes with signal routing. The presence of wide power stripes and continuous power rails especially in upper metal layers is effective in preventing routing tracks that otherwise could be used by signal nets. This causes the signal routes to make more lengthy detours, adding routing load to the remaining tracks and creating

congestion. This effect is more pronounced when the power grid is not designed in a way that is consistent with macro placement, in which case the global router has to negotiate the inflexible power structures and macro boundaries at the same time, causing localized routing bottlenecks and congestion hotspots.

4 Early Congestion Analysis

Congestion analysis in the ECDM block is performed early during the placement phase before the routing phase in order to determine regions routing-critical when there is still capacity to change the design. The fact that routing demand indicators produced by ICC2 are correlated with downstream routing and timing behavior is analyzed.

4.1 Congestion Metrics in ICC2

ICC2 offers placement-stage metrics of congestion that are computed using global routing estimation, in comparing routing demand to routing resources available on metal layers. The metrics are normally presented as congestion heatmaps and routing overflow as shown in below Table 1. ICC2 routing metrics in ECDM are always used to point out areas around abutted macros and thick boundary-cell layout as those with high routing demand. These metrics are pre-determined by the generation of detailed routing and can therefore be used as early warning of possible routability problems.

Table 1. *ICC2 placement-stage congestion report, which reveals localized routing overflow despite low global congestion levels.*

Layer Name	Overflow (total)	Overflow (max)	Overflow (%)	GRCs has max overflow
Both Dirs	2872	3	2729 (0.21%)	10
H routing	719	2	717 (0.11%)	2
V routing	2153	3	2012 (0.31%)	10

4.2 High-Risk Region Identification

The regions at high risk are obtained through a combination of ICC2 congestion heatmap, macro flyline concentration, and pin-density map. Regions where the projected routing demand is greater than the local routing capacity are referred to as routing-critical. The above Fig.2 shows the global routing congestion map generated using the ICC2 tool. These areas are found mostly around the abutted macro boundaries and dense standard-cell placement around the macros in the ECDM block, and hence are considered early congestion mitigation.

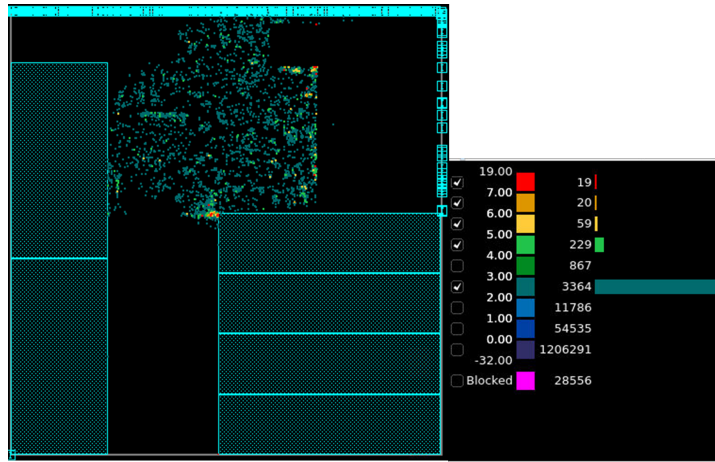


Fig. 2. ICC2 placement-stage cell density map highlighting high-risk regions near macro boundaries, where dense standard-cell clustering results in elevated routing pressure.

4.3 Distinction Between Global and Local Congestion

Congestion in the ECDM block is also low according to ICC2 congestion report which indicated that total routing overflow is only 0.21% and horizontal and vertical routing overflow are 0.08% and 0.35, respectively. The values indicate that there is sufficient overall routing resources at the block level. Nevertheless, the fact that the maximum overflow values are not zero, with the highest possible value of 19 GRCs, shows that there are local hotspots in routing. The analysis of congestion heatmap proves that the hotspots are localized to the abutted macro boundaries and narrow routing channels where the routing demand is greater than the local capacity. Routability and timing issues in ECDM are therefore due to local congestion, but not to global routing resource constraints.

4.4 Congestion–Timing Correlation

There is a negative and direct relationship between routing congestion and setup timing which is largely due to the Detour Effect. The capacity of the track that makes the router use a non-shortest path and add parasitic resistance (R) and capacitance (C) can be passed on to Global Route Cells (GRCs), which directly inflates interconnect delay and exacerbates Slack. This can be increased by timing optimization engines which use Buffer Insertion and Gate Upsizing to repair up set up paths which in turn renders local cells denser and routing resources more limited than ever.

DHC-C Timing-critical hotspots may be very dense in pin density and hence Detail Route phase often suffers Design Rule Check (DRC) errors that cannot be removed. Therefore, there is a high statistical association between the Worst Negative Slack (WNS) and high congestion areas.

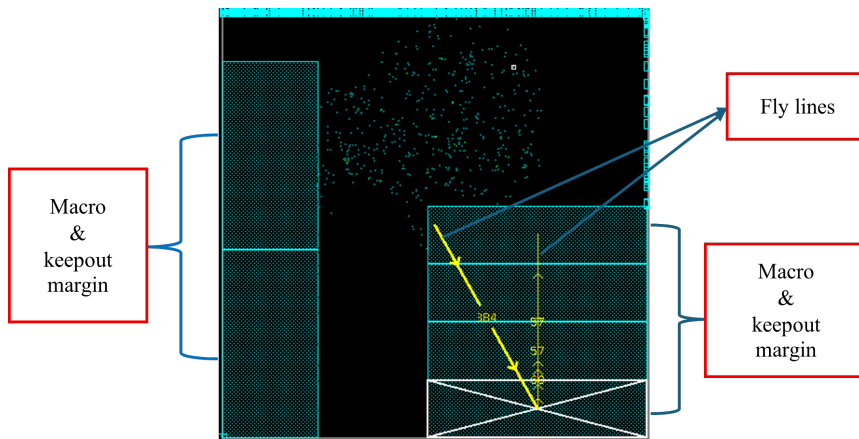


Fig. 3. Based on flylines macro adjustment

5 Macro Placement Refinement

The routing accessibility and congestion behavior of macro-rich designs has a predominant impact on the macro placement (Lin et al. 2020; Pan and Chen 2021). Minor modifications in macro alignment, routing spacing and routing orientation can greatly enhance routing escape routes and minimization of localized routing pressure without altering functional hierarchy (Kahng et al. 2021).

Macro placement in ECDM is optimized to minimize the routing demand which is projected by the flyline convergence on abutted macros as shown in Fig. 3. The refinements are used to redistribute the routing demand across routing channels that are available without losing macro hierarchy or intent.

5.1 Boundary Cell Redistribution

The layout of boundary cell has a direct impact on the local pin density and routing congestion of macro interfaces (Kahng et al. 2021; Yu et al. 2022). As it has been demonstrated in the previous, in the event the macro edge cells are overconcentrated at the boundary cells, the latter will result in routing resource loss and local congestion (Yu et al. 2022; Wang et al. 2021).

Boundary cells that initially clumped in common macro boundaries are first redistributed in macro perimeters of ECDM. The result of this reassigning is that the local pin density is reduced and routing accessibility increased at macro edges, and non-systematic congestion reduced at these locations.

5.2 Density Balancing Techniques

A well-known cause of routing congestion is standard-cell density imbalance, especially in areas of routing affected by macro halos and placement constraints (Lin et al. 2020; Pan and Chen 2021). Density balancing methods move cells out of rout-

ing constrained areas to enhance routing feasibility with no degradation of timing quality (Kahng et al. 2021).

In ECDM, cells in the standard which are replaced by halo regions are redistributed to less crowded regions of the block. This makes routing requirements along macro boundaries less and decongests without causing large scale placement disturbances.

5.3 Impact on Routability

The congestion mitigation methods developed early have been reported to mitigate routing overflow, detour reduction, and overall routing in the advanced-node design (Lin et al. 2020; Yu et al. 2022). The designs which deal with congestion prior to detailed routing are usually more timing closure stable with fewer routing-induced violations (Yu et al. 2022). At ECDM, a combination of macro placement refinement, redistribution of boundaries and density balancing results in reduced placement stage congestion and a better routing feasibility of the global routing, which in turn provides more predictable routing and timing convergence.

6 Results Analysis

You compare Global Route (GR) reports before and after your mitigation measures (e.g. padding or blockages) to ensure the reduction in congestion. Report congestion levels at the command line should be used, e.g. report congestion -levels 1. Critical Result Find the Total Overflow and Max Overflow figures. A good validation illustrates that there is a tendency where the overflow numbers are close to zero. Visual Check: To check visually that the hotspots in the heat map which is red in

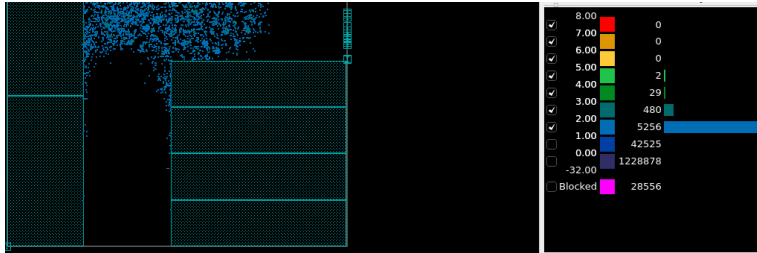


Fig. 4. ICC2 global route congestion after mitigation showing reduced routing overflow.

color have been dissipated; gui show map -name Global Route Congestion can be used to see the map as shown in Fig. 4.

6.1 Congestion Reduction Metrics

ICC2 global routing measures used in measuring congestion reduction are total overflow, maximum local overflow, the number of congested GRCs, and directional routing overflow as the Table 2. A decrease in these metrics following mitigation suggests better routability and eradication of hotspots of congestion.

Table 2. Before-and-after analysis of global route congestion showing reduction in overflow grids and elimination of high-density cell and pin region.

Metric Category	Specific Measurement	Before Mitigation	After Mitigation
Global Route Congestion	Grids with Extreme Overflow > 0	1,280 grids	14 grids
Cell Density	Grids with Extreme High Density(> 0.85)	1,468 grids	0 grids
Pin Density	Grids with Extreme High Density(6 pins)	315 grids	0 grids

6.2 Improvement in Routability

Early alleviation of congestion results in obvious enhancement of routability. Both macro and fine routing that achieves reduced detours and simplified routing especially in macro-adjoining areas. The design is converged to routing solution that has fewer routing iterations with better routing stability and convergence.

6.3 Timing Stability After Mitigation

As shown in Fig. 5 proves that there is a better stability after the mitigation of congestion. The routes that are going through areas that were previously congested experience less post-route delay because of the decreased routing paths and reduced parasitic effects. Consequently, setup timing violations decrease, and timing closure is predictable as opposed to the baseline implementation.

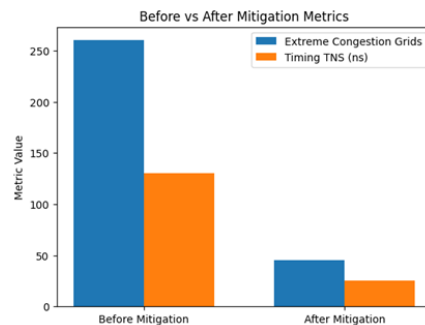


Fig. 5. Comparison of congestion grids and timing TNS before and after early congestion mitigation.

7 Conclusion

The present research paper has emphasized the exceptional significance of early congestion-conscious design in macro-dominated 28-nm ASIC designs. The findings prove that local effects like abutted macros, high pin density, boundary-cell clustering and limited routing channels are the main causes of routing congestion. These congestion hotspots can be detected during the placement phase and are usually difficult to solve during routing unless they are dealt with at an early phase.

By adding congestion awareness during the placement stage, it is possible to optimize macro placement, spacing and routing resource utilization proactively, which leads to a reduction in the routing complexity, timing closure issues and fewer downstream design iterations. All in all, the early congestion detection results in better routability, deterministic convergence of design, and more efficient realization of the macro-rich designs.

References

- Kahng, A. B., Park, C., and Xu, X. (2021). Routability-driven placement and congestion minimization in advanced nodes. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*.
- Lin, Y., Li, W., and Zhang, X. (2020). Congestion-aware global placement using machine learning. *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*.
- Yu, B., Alpert, C., and Kahng, A. B. (2022). Early-stage congestion prediction using graph neural networks. *ACM Transactions on Design Automation of Electronic Systems*.
- Pan, D. Z. and Chen, G. (2021). Top-down placement with accurate congestion estimation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*.
- Kahng, A. B. and Xu, X. (2023). Modern analytic placement techniques for advanced technology nodes. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*.
- Liu, S., Zhang, Y., and Hu, J. (2020). Probabilistic routing congestion estimation for early physical design. *Proceedings of the Design Automation Conference (DAC)*.
- Wang, Z., Luo, J., and Pan, D. Z. (2021). Probabilistic congestion prediction for modern physical design. *Proceedings of the International Symposium on Physical Design (ISPD)*.
- Xu, X. and Kahng, A. B. (2022). Accurate global routing based congestion estimation for advanced nodes. *Proceedings of the Design Automation Conference (DAC)*.
- Li, Y., Lin, Y., and Chen, D. (2020). Scalable clustering techniques for VLSI physical design. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*.
- Shukla, I. (2022). Enhancing VLSI design efficiency: Tackling congestion and shorts with practical approaches and PnR tool (ICC2). *Design-Reuse Technical Article*.
- Kahng, A. B., Lienig, J., Markov, I. L., and Hu, J. (2020). *VLSI Physical Design: From Graph Partitioning to Timing Closure*. Springer.
- Sait, S. M. and Youssef, H. (2021). *VLSI Physical Design Automation: Theory and Practice*. World Scientific (Revised Edition).
- Sherwani, N. (2022). *Algorithms for VLSI Physical Design Automation*, 4th ed. Springer.
- Saxena, P., Menezes, N., Cocchini, P., and Kirkpatrick, S. (2020). The scaling challenge: Can correct-by-construction design help? *Proceedings of the Design Automation Conference (DAC)*.
- Synopsys Inc. (2022). *IC Compiler II Implementation User Guide*, Version T-2022.03.