# Inductance-aware Clock Network Synthesis Considering Hierarchical Interconnects in 3D ICs

Jindong Zhou
ShanghaiTech University
shanghai, China
Shanghai Institute of Microsystem and Information
Technology, Chinese Academy of Sciences
shanghai, China
University of Chinese Academy of Sciences
beijing, China
zhoujd@shanghaitech.edu.cn

Zi'Ang Ge
ShanghaiTech University
shanghai, China
geza2022@shanghaitech.edu.cn

Chenbo Xi
ShanghaiTech University
shanghai, China
xichb2023@shanghaitech.edu.cn

Pingqiang Zhou
ShanghaiTech University
shanghai, China
zhoupq@shanghaitech.edu.cn

## Abstract

Designing a novel clock network is facing increasing challenges, particularly in Three Dimensional Integrated Circuits (3D ICs) where the clock signal traverses a complex interconnect path consisting of bumps, Through Silicon Vias (TSVs) and multi-layer metal wires. During the analysis in 3D Clock Network Synthesis (CNS) process, previous works overlook 1) the impact of inductance on delay and 2) the hierarchical interconnect structures in 3D ICs, which can result in significant deviations between the ideal delay-balanced design produced by algorithms and real-world scenarios. Consequently, clock skew, which represents the balance of the clock network, experiences considerable increases and finally affects system performance. In this work, the inductance of every interconnect hierarchy is considered and the RLC delay model is applied in the analysis to present different interconnect delay characteristics in a 3D clock network. Some algorithms are also developed and modified in 3D CNS to adapt to the proposed models in order to keep the skew as small as possible. Then an inductance-aware strategy is applied in topology generation to further improve the 3D clock network. The results demonstrate comparable quality with prior works with regard to TSV numbers and wire length, etc. The proposed method outperforms prior works – it can catch the characteristics in 3D CNS and reduce the average clock skew from 364$ps$ to 90$ps$ on 3D-ISPD09 benchmarks. Besides, the skew can be further reduced by the improved topology generation method.

## CCS Concepts

• **Hardware → 3D integrated circuits**; **Clock-network synthesis**.

## Keywords

3D-IC, clock network synthesis, elmore delay, interconnect model, parasitic inductance

## 1 Introduction

Three Dimensional (3D) Integration technologies have already been proposed and applied in commercial products, like '3D V-cache' from *AMD* [29]. Through Silicon Vias (TSVs) and chip-stacking bumps are key cross-tier interconnect components for vertical connections within a single chip package [11, 13]. Composed of simple interconnect wires and buffers, the clock distribution network consumes not only substantial routing and driver resources, but also a significant fraction of power in a synchronous digital system due to continuous clock signal switching [19]. Therefore, the planning and physical design of the on-chip clock network, completed through Clock Network Synthesis (CNS) step in chip design flow [25], are crucial for designing a high-performance and low-power 3D chip.

The basic constraint for CNS is to find a balanced clock distribution network that can minimize clock skew among all sinks [3, 6]. In addition, given the limited routing resources, CNS algorithms also minimize the wirelength and power consumption of the clock network [5, 6]. However, transitioning from 2D to 3D chips introduces new challenges and necessitates additional constraints. As shown in Figure 1, the clock sinks of a 3D chip are not gathered on one die but are distributed over multiple dies sharing a single clock network, which means the physical transmission path of a
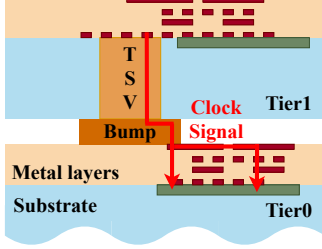
**Figure 1: An example of a clock signal transmission path in the TSV-based face to back 3D IC interconnect hierarchy: one tier → TSV → bump → global metal → local metal layers in another tier.**
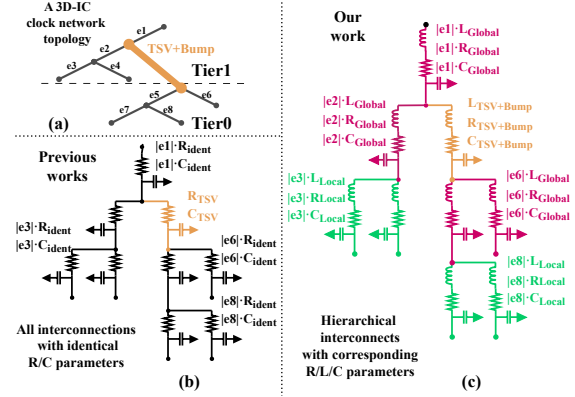


**Figure 2: (a) A 3D clock network topology: leaves are clock sinks and branches represent the abstract connections. (b)(c) Model comparisons of 3D clock network.**
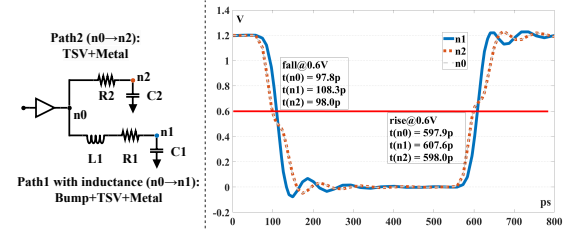


**Figure 3: A simplified case about the inductance effect in 3D interconnect.** $Delay_{n0-n2}(RC) \approx 0.2ps$, $Delay_{n0-n1}(RLC) \approx 10ps$. **(Parameters are from Section 4.1. Metal level is 'Global' and the length is 50um.** $R1 = 0.587\Omega$, $C1 = 30.66fF$, $L1 = 0.609nH$, $R2 = 0.585\Omega$, $C2 = 29.63fF$.**)**

clock signal becomes complex with the existence of bumps, TSVs and multi-level metal wires [11]. Therefore, 3D CNS becomes more challenging than that in 2D case.

In the literature, several pioneering works have proposed complete CNS flow for 3D ICs with the objective of minimizing the clock skew. Method of Means and Medians (MMM) [31] follows top-down topology generation principle, and Kim [9] proposes a bottom-up flow based on Nearest Neighbor (NN) selection to recursively select pairs of sub-trees with the lowest merging cost and then merge them with zero skew through the Deferred-Merge Embedding (DME [3]) method. NN-based strategy has shown better performance than MMM in terms of clock skew and routing resources [9]. Due to its effectiveness, in the past decade, NN-based method has been actively adopted in 3D CNS works [10, 12, 17, 27, 30]. Besides, researchers have looked into other issues in the 3D CNS topic like electromigration [14], low-power [15], buffer insertion [21], etc.

However, previous works overlook two crucial aspects:

- *The impact of inductance on 3D interconnect delay.* The inductance impact of TSVs has been discussed by researchers on 3D IC design topics such as delay analysis [2] and power delivery network design [20]. In CNS flow, one critical step is the delay calculation of each net in the clock network. However, inductance is ignored by previous works on 3D CNS. Instead, RC model (see Figure 2(b)) is used for delay calculation due to its simplicity and efficiency. Also, to the best of our knowledge, the latest EDA tool *Integrity 3D-IC* [1] is still not mature enough to support inductance-aware 3D CNS in physical design step. However, bumps, TSVs and high-level metal layers in advanced 3D ICs typically possess relatively large sizes (in *um*), resulting in significant parasitic inductance. It is accepted that inductance can have large impact on signal delays [7, 26] in long interconnects, and it is critical to carefully choose an appropriate interconnection model for RLC delay analysis in CNS [4]. Furthermore, advanced circuits operate at high frequencies (*GHz*), exacerbating the impact of inductance on delay. As demonstrated by the example in Figure 3, even if the resistance and capacitance are almost identical, the inductance induced by 3D vertical interconnect components brings considerable impact on delay. Then this delay difference can accumulate with the increasing depth of clock network. Besides, after introducing inductance, solving delay balance equation becomes difficult. Therefore, the inductance effects must be considered and efficient algorithms are required.

- *The hierarchical interconnect structures in 3D ICs.* As shown in Figure 1, the clock transmission paths consist of heterogeneous components including bumps and TSVs (for cross-tier transmission) and hierarchical metal layers (inside one tier). However, metal layers with different width/spacing/thickness exhibit different parasitic parameters, but they are often indiscriminately treated as a simplified and identical interconnect type (see Figure 2(b)) in all prior works. In view of that the practice scenario deviates from ideal assumptions in those works and that the vertical interconnect and global metal with large sizes bring non-negligible inductance, the delay analysis of 3D clock networks must be more accurate and meticulous.

In this work, we consider the inductance impact on clock delay, establish more practical models and optimize the algorithms in 3D CNS to further improve the quality of the 3D clock network. Our contributions can be summarized as:

- We consider the inductance components in the 3D interconnect hierarchy and adopt the RLC delay models in analysis. An approximation algorithm is implemented to efficiently find the position of the merging node.
- We develop an inductance-aware topology generation method to achieve better topologies and control vertical interconnect consumptions.
- We introduce a hierarchical interconnect model for the analysis of 3D clock networks, including the bumps and TSVs, as well as distinct metal interconnects.

- Our approaches are evaluated on modified ISPD09 benchmark set. Results show that our method outperforms previous NN-based 3D CNS flows – it can reduce the average clock skew from 364*ps* to 90*ps*, with comparable quality in routing resources and power. The clock topology can be further optimized for smaller clock skew.

## 2 Preliminary

Generally, the 3D CNS process comprises a sequence of steps [25] (see Figure 4): (1) abstract clock topology generation; (2) layer embedding of internal nodes; (3) network routing; and (4) buffer insertion. The topology generation constructs an abstract interconnect topology to connect all clock sinks. Note that in this topic it only considers a clock network with one clock source and numerous sinks distributed at multiple tiers all over the 3D IC. The layer embedding step assigns a 3D chip tier ID to each internal node of the topology. Then, clock routing determines the precise position of each internal node to balance the delay between its two branches. Buffer insertion is often performed simultaneously to minimize the total wire length and buffer number. Some basics of the related algorithms [3, 5, 9] are not elaborated on in this article.

The objective of CNS is to ensure a uniform distribution of clock signals throughout the entire chip, that is, zero clock skew. Skew is the difference in arrival times of a clock signal at two clock sinks. Typically, the skew of one design is the maximal value among all sink pairs. In practical scenario, it cannot achieve the ideal zero. Hence, minimizing the maximal skew is preferable. In addition, other metrics are also used to assess the quality of a 3D clock network design, such as total wire length, TSV count, buffer count and power consumption.
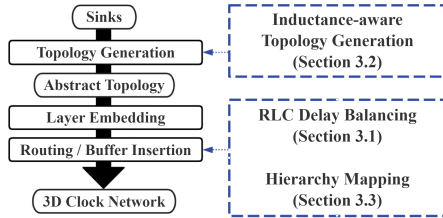


**Figure 4: The 3D clock network synthesis flow and our proposed approaches in corresponding steps.**

## 3 Proposed Approachs

In this work, firstly, we use the RLC delay model, instead of the *Elmore* delay model, in order to get close to practical situation in our 3D CNS flow. Considering that it is too complex to solve the RLC delay balance equation directly, we propose an efficient algorithm to find a proper position of the merging node in the routing step. Secondly, in the topology generation step, we design an inductance-aware strategy to generate an optimized topology. It also has the ability to control the routing resource consumptions. Thirdly, we employ hierarchical interconnect models and utilize the corresponding electrical parameters for the nodes and their interconnects in the 3D clock network.

### 3.1 RLC Delay Balancing in 3D CNS

In 3D CNS, once the layer embedding step assigns the tier ID of nodes within the clock tree, the routing step using DME [3] method
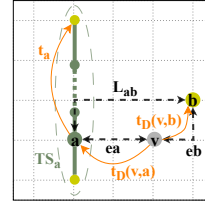


- nodes $a$, $b$: the child nodes of the merging node $v$ in a clock tree topology.
- $|e_a|$, $|e_b|$, $L_{ab}$: the 2D Manhattan distance between node $v\&a$, $v\&b$ and $a\&b$, thus $|e_a| + |e_b| = L_{ab}$.
- $t_D(v, a)$, $t_D(v, b)$: the interconnect delays for the signal to propagate from node $v$ to nodes $a$ and $b$.
- $t_a$, $t_b$: the delay from the root node $a$ and $b$ to the leaf node (i.e. clock sink) in the sub-trees $TS_a$ and $TS_b$.
- $t_DWire$: the propagation delay of the metal wire between two nodes based on the 2D Manhattan distance.
- $t_DTSV$: the delay of a vertical interconnect path.

**Figure 5: Notations in the CNS problem. A sub-tree can be a sink like node $b$ in the illustration and $t_b = 0$.**

determines the coordinates of every node to balance the delay and minimize clock skew from bottom to up in the topology. At each step of the DME algorithm, the positions of node $a$, $b$ have been determined, then the sub-problem is to determine the position of the merging node $v$ based on the edge $e_a$, $e_b$. It is shown in Figure 5 with notations listed.

The total signal delay from every merging node $v$ through node $a$ to clock sinks is: $T_a = t_a + t_D(v, a)$. If nodes $v$ and $a$ do not belong to the same chip tier, TSVs should be inserted and $t_D(v, a)$ becomes $t_DTSVs + t_DWire(v, a)$. Note that, for convenience, the term 'TSV' in this article refers to a combination of a TSV and a bump, which are treated as a lumped RLC model in delay calculation. To find $t_D(v, a)$ and $t_D(v, b)$ with RLC models exhibiting non-monotonic responses, we adopt an efficient second-order approximation method [7] to characterize the signal propagation delay:

$$t_{pdi} = (1.047e^{-\frac{\zeta_i}{0.85}})/\omega_{ni} + 0.695 \sum_k C_k R_{ik}, \qquad (1)$$

$$\zeta_i = \frac{1}{2} \frac{\sum_k C_k R_{ik}}{\sqrt{\sum_k C_k L_{ik}}}, \quad \omega_{ni} = \frac{1}{\sqrt{\sum_k C_k L_{ik}}}. \qquad (2)$$

$k$ is the index of downstream nodes of node $i$. Note that a small $\zeta_i$ reflects a relatively large inductance.

So the sub-problem is to merge the two sub-trees while satisfying zero clock skew, which is to make the following equation hold:

$$t_a + t_D(v, a) = t_b + t_D(v, b) \qquad (3)$$

To find the position of node $v$ (merging segment), we need to solve equation (3) to get $|e_a|$ and $|e_b|$. However, it is extremely difficult since the origin RLC delay formula contains exponential and square root calculations and TSV delay may exist. Luckily, we observe that the primary component of the interconnection delay is the RC delay. So instead of solving equation (3) directly, we use an approximation strategy to find a proper position of node $v$ accurately and efficiently. The schematic diagram of the method is illustrated in Figure 6. The *Elmore* delay model is first adopted to get a solution ($v_0$) because it is just a simple quadratic equation which



**Figure 6: To find the position of merging node $v$ to balance the RLC delay.**
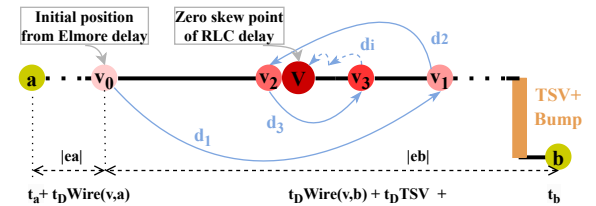
is easy to solve. Then based on this initial position or the so-called merging segment of node $v$, $T_a$ and $T_b$ are calculated in RLC delay model. If $T_a < T_b$, then increase $|e_a|$ by a displacement $d_1 = |e_b|/2$ to find a new solution node $v_1$. This process iterates with a dynamic displacement distance $d_i$ until the delay difference between two branches is under an extremely small threshold ($\theta$). Following the idea of the binary search, the direction of the displacement is based on the sign of $T_a - T_b$ and the distance is halved with iterations until 1 $nm$. The pseudo code is shown in Algorithm 1.

---

**Algorithm 1** Tuning $|e_a|,|e_b|$ for RLC delay balance.

**Input:** $|e_a|,|e_b|$ solutions based on *Elmore* delay model
**Output:** $|e_a|,|e_b|$ for balancing the RLC delay
1: Calculate $t_D(v, a)$ and $t_D(v, b)$ using RLC delay model;
2: **while** $|T_a - T_b| > \theta$ **do**
3:   **if** $T_a < T_b$ **then**
4:     Tune $|e_a|,|e_b|$ by a dynamic displacement;
5:     **if** $|e_b| = 0$ **then**
6:       Increase $|e_a|$ only to enlarge the *TRR* of node $a$;
7:   **else**
8:     Operate in opposite ways;
9:   Re-calculate $t_D(v, a)$, $t_D(v, b)$ with updated $|e_a|$, $|e_b|$;
  *: Details are described in the origin DME [3] algorithm.

---

## 3.2 Inductance-aware Topology Generation

As mentioned in Section 1, the widely adopted topology generation strategy is a bottom-up approach: nearest neighbor selection because it can generate 3D clock networks with less resource and power consumptions and smaller skew. Firstly in NN-based methods, the cost of edges in a nearest neighbor graph of all sub-trees and sinks are calculated and sorted in increasing order. Then the first pair in the order is merged to form a new node and the cost of merging this new node with all other remaining nodes should be updated. The process repeats until one tree is left. Prior works use capacitance value when estimating the merging cost. However, capacitance alone can not accurately reflect the practical RLC delay, leading to a sub-optimal clock topology. Incorporating inductance information into the merging cost can generate a clock topology that better fits the RLC practice and leads to a better 3D clock network. In fact, the quality of the generated abstract clock topology can significantly affect the resources consumed. Figure 7 shows an example. While it just contains eight sinks, it underscores the vast solution space in 3D CNS problem.

Obviously, there is potential for optimization in 3D clock network by generating an appropriate topology, especially when considering the parasitic inductance. We estimate the merging cost with RLC delay information by the following proposed formulas:

$$C(a, b) = t_{est}W(a, b) + \alpha t_D TSVs(a, b) + Max(t_a, t_b), \quad (4)$$

$$t_{est}W(a, b) = \sqrt{c'l'} + 0.695r'c', \quad (5)$$

where the last term in formula (4) is the downstream delay set to merge sub-trees with smaller delay earlier. The cost of merging node $a$ and $b$ is based on the estimated metal wire delay and whether there exists TSVs between them. Formula (5) is simplified from the origin RLC delay formula (1)&(2). The $(r', l', c')$ are only based on $L_{ab}$ and the parasitic parameters are the ones of the 'Local' metal

level temporarily (described in Section 3.3). Besides, one weight parameter $\alpha$ is added to represent the importance of the TSV delay. For instance, if setting a high $\alpha$, which enlarges the merging cost of the node pairs requiring TSVs, the algorithm will not tend to choose these pairs. As a result, the final clock network consumes fewer TSVs but potentially more metal wire resources.

Our strategy, by catching inductance factors and determining the 3D clock topology, has the ability to constrain routing resources, particularly TSVs, and has the potential to improve the final clock network performance in terms of skew.

## 3.3 Hierarchy Mapping Considering Inductance

As shown in Figure 1, clock signals can traverse through bumps and TSVs across multiple tiers. For a single die, the clock signals propagate throughout the die via metal interconnects. In inductance-aware delay analysis, explicit definition of electrical parameters is required for each interconnect hierarchy, especially for the bumps, TSVs and global metal wires (see Figure 2(c)) because they induce relatively large inductance than their resistance and capacitance. Particularly in the routing step, the clock topology needs to be mapped to a certain interconnect hierarchy.

The generated abstract clock topology consists of multiple layers from the sub-tree root to the leaves in one tier. Those topology layers can be divided into several portions. Meanwhile, the metal layers can be broadly categorized into two levels: global and local according to their significant size differences. We map the top topology portions to 'Global' metal level and the bottom ones to 'Local'. Thus, we can use RLC delay model to calculate the delay through TSVs, bumps and high level metal wires for the accurate inductance-aware delay, while use *Elmore* delay model to calculate the delay through lower metal wires efficiently because the inductance effect can be neglected in this case. The proposed strategy takes into account the size differences as well as parasitic characteristic differences among metal layers without delving into the detailed parameters of each individual metal layer. The number of the portions and the metal levels is not fixed and can be adjusted according to the specific process used. If needed for an extremely accurate analysis, our method can be easily extended to every metal layer in detail together with complete RLC delay calculations.
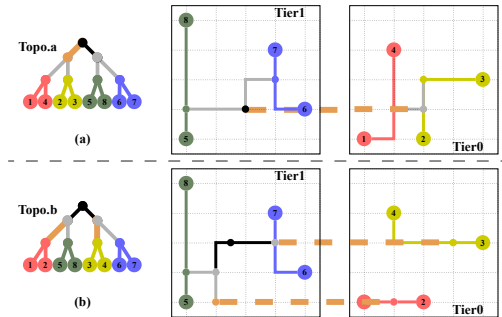


**Figure 7: An eight-sink clock network example with two different abstract clock topologies. In final clock networks, *Case.a* and *Case.b* leverage 1 and 2 TSVs while consume 21 and 18 units of wire, respectively.**
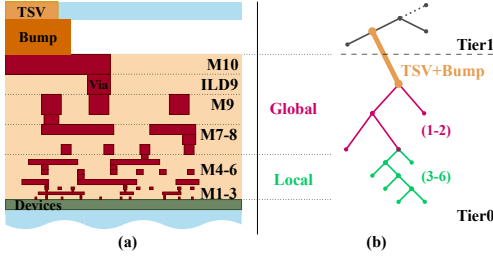
**Figure 8: (a) Cross section schematic of metal layers of $45nm$ process; (b) Illustration of clock network topology mapping.**

Taking the $45nm$ case (*FreePDK45* [16]) in Figure 8(a) as an example, there are ten metal layers, which can be classified into two levels: 'Global' (M7-10) and 'Local' (M1-6). Supposing there is an abstract clock topology shown in Figure 8(b). Some nodes in 'Tier0' form a tree with six topology layers, and the clock signal is transmitted into this sub-tree. The first two topology layer is mapped to 'Global' and the others are 'Local' (presented by different colors).

## 4  Experiment Results

### 4.1  Experimental Setup

Our 3D CNS algorithm is implemented using *C++* and runs on a Linux OS with a $3.4GHz$ Intel 4-core processor. Following the related works [9, 14, 15], we evaluate the performance of our 3D CNS flow on eleven benchmarks adapted from the ISPD09 clock network synthesis contest [24]. Each benchmark features a unique die size and varies in the number of clock sinks, ranging from 91 to 440. Each sink has its own location and load capacitance information. Since these benchmarks are originally designed for 2D CNS, we transform them into 2-layer 3D circuits as done in the references [9, 14, 15]: The die footprint is reduced by a factor of $\sqrt{2}$, and the tier IDs of sinks are assigned randomly. Note that the location and tier ID of the sinks can also be determined by designers in practice, and we believe our method also works in that case. It should also be pointed out that the random assignment of sinks in the experiments also reflects the trend of full logic on logic instance level 3D integration, which has already been discussed in prior academic works [13, 25] and has been proposed by AMD [22]. The benchmarks exclusively focus on the clock network, so all evaluation metrics are only about the corresponding clock network.

In our work, the $45nm$ technology process serves as the illustrative case and the resistance, capacitance and inductance values of the metal levels are calculated based on the analytical models [18, 28] and the interconnect sizes are from *FreePDK45* [16], which are listed in Table 1. Note that if a wire is longer than $50um$, it should be divided into multiple segments in delay analysis in case of the inaccuracy of the lumped RLC circuit model.

The parasitic parameters of bumps and TSVs are obtained from [8, 9] and the resistance, capacitance and inductance per unit of bump

/ TSV are $2/35m\Omega$, $1.03/15.48fF$ and $1.39/13.83pH$. The buffer characteristics are: input capacitance: $35fF$, output capacitance: $80fF$ and output driving resistance: $61.2\Omega$. The supply voltage is set to $1.2V$. All metrics of the 3D clock network are evaluated by SPICE simulation. All evaluations in this work are under the same aforementioned practical RLC delay model and parameters to maintain the consistency and comparability of the results.

### 4.2  Clock Skew Reduction by the Proposed 3D CNS Flow

The baseline results (***Base***) are evaluated according to the NN-based flow, which is widely adopted [9, 10, 12, 17, 27, 30]. The resource consumption results are basically consistent with the original literature, except for power because of the different parameters used. To demonstrate the effectiveness of our proposed approaches, we compare the baseline with our three progressive 3D CNS flows: 1) ***L*** is based on the origin flow considering parasitic inductance, 2) ***L&H.*** further considers the proposed hierarchical interconnect model, and 3) ***L,H.&TO.*** is our complete flow with simple ($\alpha = 1$) inductance-aware topology generation. The results are summarized in Table 2. The last column *ratio* represents the average of all ratio values for the respective benchmarks compared to the baseline.

Regarding the identical abstract network topology generated by the same algorithm in the first three flows, when considering the parasitic inductance in the algorithms (method: ***L***), the clock skew can be reduced by nearly 30%, which shows the necessity of considering the parasitic inductance in 3D CNS algorithms to help balance the clock delay. When the hierarchical interconnect model is applied (method: ***L&H.***), it can effectively reduce the average clock skew from the baseline of $364ps$ to $98ps$, which proves the importance of the proposed inductance-aware hierarchical interconnect model in the analysis. The results show an increase of about 2.02% in wire length and an increase of 1.69% in buffer number, which finally causes a 1.69% increase in power. The baseline (method: ***Base***) encounters unfavorable skew results due to the absence of models that cover parasitic inductance and hierarchical interconnect information. Thus, the imbalance delay errors accumulate remarkably in the bottom-up algorithm and ultimately result in unsatisfactory skew results in the evaluation. Our flows can capture these characteristics, yielding more balanced clock networks.

### 4.3  Further Skew Reduction by Inductance-aware Topology

As shown in the last row of each metric in Table 2, by applying the proposed inductance-aware topology generation strategy (Section 3.2), our complete 3D CNS flow (method: ***L,H.&TO.***) can further reduce the average skew to $90ps$ with almost the same routing resource and power consumption albeit using a few more TSVs. The quality of the abstract topology generated in the first step in 3D CNS is critical to the final clock network metrics, which means it is better to optimize the 3D clock network design in early stages and consider some factors like the non-negligible inductance information in advance. The results of *fnb1* and *fnb2* are greatly affected in this method that the wire length decreases while the TSV number increases substantially.

**Table 1: Parameters of metal levels in $45nm$.**

| Metal Level | Wid.&Spac. ($nm$) | Thick. ($nm$) | ILD T. ($nm$) | Resist. ($\Omega/um$) | Capacit. ($fF/um$) | Induct. ($pH/um$) |
|---|---|---|---|---|---|---|
| *Global* | 800 | 2000 | 2000 | 0.011 | 0.283 | 11.883 |
| *Local* | 70 | 140 | 120 | 1.714 | 0.267 | 12.128 |

**Table 2: Performance evaluations on eleven ISPD09 benchmarks modified for 3D CNS under the RLC delay model.**

| | Methods | f11 | f12 | f21 | f22 | f31 | f32 | f33 | f34 | f35 | fnb1 | fnb2 | ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Skew** (ps) | Base | 321 | 274 | 408 | 218 | 555 | 659 | 487 | 434 | 449 | 79 | 123 | 1.000 |
| | L | 204 | 204 | 268 | 180 | 375 | 563 | 316 | 398 | 308 | 44 | 71 | 0.707 |
| | L&H. | 88 | 109 | 81 | 66 | 116 | 153 | 191 | 79 | 92 | 42 | 69 | 0.317 |
| | L,H.&TO. | 66 | 58 | 77 | 75 | 138 | 142 | 87 | 123 | 102 | 58 | 68 | 0.308 |
| **Wire Length** (mm) | Base | 126.9 | 112.6 | 131.6 | 73.4 | 268.5 | 203.2 | 204.9 | 171.0 | 194.8 | 28.8 | 71.0 | 1.000 |
| | L | 126.8 | 113.6 | 130.5 | 72.7 | 270.7 | 205.5 | 207.3 | 169.9 | 196.6 | 28.5 | 72.7 | 1.003 |
| | L&H. | 127.7 | 115.2 | 132.5 | 73.2 | 274.2 | 217.7 | 209.6 | 172.6 | 203.7 | 28.5 | 72.8 | 1.020 |
| | L,H.&TO. | 132.8 | 118.1 | 136.1 | 78.9 | 276.9 | 211.8 | 212.5 | 185.4 | 201.1 | 27.9 | 72.3 | 1.038 |
| **#TSVs** | Base | 33 | 36 | 31 | 27 | 84 | 61 | 65 | 47 | 61 | 53 | 109 | 1.000 |
| | L | 33 | 36 | 31 | 27 | 84 | 61 | 65 | 47 | 61 | 53 | 109 | 1.000 |
| | L&H. | 33 | 36 | 31 | 27 | 84 | 61 | 65 | 47 | 61 | 53 | 109 | 1.000 |
| | L,H.&TO. | 41 | 41 | 35 | 27 | 90 | 65 | 72 | 48 | 66 | 109 | 140 | 1.199 |
| **#Bufs** | Base | 163 | 145 | 155 | 105 | 329 | 251 | 255 | 199 | 233 | 81 | 159 | 1.000 |
| | L | 165 | 151 | 155 | 103 | 331 | 249 | 251 | 199 | 239 | 79 | 161 | 1.003 |
| | L&H. | 167 | 151 | 159 | 107 | 331 | 251 | 257 | 201 | 245 | 79 | 163 | 1.017 |
| | L,H.&TO. | 159 | 147 | 161 | 111 | 333 | 257 | 259 | 205 | 245 | 81 | 159 | 1.020 |
| **Power** (mW) | Base | 85.4 | 76.8 | 85.2 | 53.2 | 178.0 | 135.6 | 137.0 | 111.8 | 128.0 | 34.8 | 68.6 | 1.000 |
| | L | 85.8 | 78.2 | 84.6 | 52.6 | 179.0 | 135.4 | 137.0 | 111.4 | 130.0 | 34.4 | 69.4 | 1.002 |
| | L&H. | 86.8 | 78.8 | 86.4 | 53.8 | 180.0 | 140.4 | 138.8 | 112.8 | 134.0 | 34.4 | 69.8 | 1.017 |
| | L,H.&TO. | 87.2 | 79.2 | 88.4 | 56.6 | 181.8 | 139.4 | 140.4 | 118.8 | 132.8 | 35.8 | 69.6 | 1.034 |
| **Max Slew** (ps) | Base | 476 | 485 | 668 | 458 | 705 | 2163 | 643 | 917 | 680 | 211 | 298 | 1.000 |
| | L,H.&TO. | 267 | 222 | 312 | 431 | 443 | 456 | 384 | 265 | 387 | 197 | 211 | 0.578 |
| **Max Frequency** (GHz) | Base | 0.42 | 0.41 | 0.30 | 0.44 | 0.28 | 0.09 | 0.31 | 0.22 | 0.29 | 0.95 | 0.67 | 1.000 |
| | L,H.&TO. | 0.75 | 0.90 | 0.64 | 0.46 | 0.45 | 0.44 | 0.52 | 0.75 | 0.52 | 1.02 | 0.95 | 2.080 |
| **Latency** (ns) | Base | 1.35 | 1.32 | 1.49 | 1.10 | 2.07 | 2.14 | 1.84 | 1.89 | 1.72 | 0.48 | 0.75 | 1.000 |
| | L,H.&TO. | 1.32 | 1.27 | 1.43 | 1.08 | 1.89 | 1.92 | 1.72 | 1.82 | 1.71 | 0.49 | 0.73 | 0.962 |

However, the skew results of some benchmarks (like *f35* and *fnb1*) fall short of expectations because they acquire the sink and TSV densities several tens of times greater than others. More sinks enlarge the network topology depth and more TSVs enlarge the asymmetry of the clock networks, so the delay is different from the ideal balance and errors accumulate [7]. To solve this problem, one strategy is to adjust the abstract topology to reduce the TSV usage to mitigate the asymmetry. Increasing the weight ($\alpha$) in Formula (4) can achieve this goal (Section 3.2). The improved results of five unsatisfying cases (using method: **L,H.&TO.** with $\alpha = 1$) by tuning $\alpha$ are shown in Table 3. *Ratio* represents the average of all ratio values compared to the simple one.Based on better clock topologies, the skew can be reduced from an average of 99*ps* to 72*ps*. At the same time, the consumption of routing resources also decreases, in particular by a 10% reduction in TSV usage. These results as well as the ones in Table 2 prove the effectiveness of the inductance-aware clock topology generation strategy.

By choosing proper $\alpha$, the 3D clock topology has the potential to be optimized. Different values of $\alpha$ in Equation (4) are set to generate different topologies to investigate the trade-offs of routing resources and to evaluate the clock network performances. The benchmark 3D-ISPD09-*fnb1* is chosen and the results are listed in Table 4. The results are consistent with the analysis in Section 3.2.

**Table 3: Performance evaluations on unsatisfactory cases.**

| | f22 | f31 | f34 | f35 | fnb1 | ratio |
|---|---|---|---|---|---|---|
| **tuned $\alpha$** | 15k | 50k | 64k | 45k | 200 | - |
| **Skew** (ps) | 54 | 104 | 85 | 67 | 53 | 0.747 |
| **Wire L.** (mm) | 77.0 | 272.0 | 176.6 | 198.8 | 29.2 | 0.989 |
| **#TSVs** | 26 | 80 | 42 | 56 | 98 | 0.895 |
| **#Bufs** | 109 | 329 | 201 | 237 | 85 | 0.993 |
| **P.** (mW) | 55.2 | 178.6 | 113.8 | 130.8 | 36.8 | 0.987 |

If enlarging the weight ($\alpha$) for TSV delay in the merging cost, the number of TSVs can decrease until one with an increase of wire length, which is a kind of routing resource utilization trade-off. The skew results can validate the previous analysis that fewer TSVs can mitigate the asymmetry of 3D clock networks and lead to smaller clock skew. Besides, the parameter $\alpha$ offers control over the number of TSVs used. Previous topology generation algorithm can not effectively choose where to insert the TSV when inductance influences exist, probably consuming the TSV budget too early to find an optimal topology. From the perspective of TSV resource constraints and 3D clock network performance, our strategy provides preferable solutions in topology generation process considering inductance.

Since ISPD09 benchmarks have a maximum of 440 sinks, to further validate the generality of our proposed 3D CNS flow, three cases with far more sinks from the ISPD10 benchmark [23] are adapted and modified to 3D scenarios similarly. As listed in Table 5, the skew results being consistent with former ones prove the scalability of our entire proposed 3D CNS flow. In addition, our flow incurs only a slight increase in runtime compared to the baseline when we include inductance in the delay calculation.

## 4.4 Slew and Latency Reduction

Clock slew and latency are two other crucial metrics for clock network. The slew time of each sink node in a given circuit is

**Table 4: Performance evaluations on different clock topologies of 3D-ISPD09-*fnb1* with varying parameter $\alpha$.**

| $\alpha$ | 200 | 800 | 2k | 10k | 40k | 200k |
|---|---|---|---|---|---|---|
| **Skew** (ps) | 53 | 65 | 61 | 56 | 49 | 48 |
| **Wire L.** (mm) | 29.2 | 28.9 | 30.2 | 34.9 | 35.6 | 38.2 |
| **#TSVs** | 98 | 76 | 53 | 16 | 6 | 1 |

**Table 5: Evaluations on benchmarks with more sinks.**

|  | Methods | 09fnb2 | 10c06 | 10c08 | 10c02 |
|---|---|---|---|---|---|
| #Sink | - | 440 | 981 | 1134 | 2249 |
| Skew | Base | 123 | 69 | 87 | 454 |
| (ps) | L,H.&TO. | 68 | 60 | 69 | 78 |
| Wire L. | Base | 71.0 | 44.1 | 50.8 | 537.3 |
| (mm) | L,H.&TO. | 72.3 | 36.7 | 42.8 | 532.7 |
| #TSVs | Base | 109 | 55 | 82 | 621 |
|  | L,H.&TO. | 140 | 325 | 359 | 759 |
| #Bufs | Base | 159 | 205 | 239 | 1393 |
|  | L,H.&TO. | 159 | 225 | 251 | 1409 |
| Power | Base | 68.6 | 68.8 | 80.2 | 510.1 |
| (mW) | L,H.&TO. | 69.6 | 75.3 | 85.3 | 514.1 |
| Runtime | Base | 2.4 | 2.6 | 3 | 5.6 |
| (s) | L,H.&TO. | 2.4 | 2.8 | 3.3 | 6.7 |

evaluated across all benchmarks and the maximum slew values are recorded. Also, based on the constraint specified in ISPD09 contest that the slew time should not exceed 20% of the clock period, the theoretical maximum operating frequency of each clock network is calculated. Previous works seldom discuss the slew performance of 3D clock networks, as the constraint can be easily met when only considering parasitic resistance and capacitance. However, the incorporation of parasitic inductance worsens the slew of the clock signal, particularly in 3D ICs where large inductance is presented. As shown in Table 2, the maximum slew time is reduced by 42% compared to the baseline. Thus the clock network generated can operate at about twice the speed. Our proposed flow has the ability to capture the inductance characteristics of the 3D clock networks and produces better buffer assignments, thereby reducing the slew. Besides, clock latency is slightly reduced by 3.8% on average due to improved buffer assignments.

## 5 Conclusion

This work introduces an inductance-aware clock network synthesis flow for 3D ICs with the hierarchical 3D interconnects. The flow incorporates parasitic inductance, utilizes the RLC delay model, employs a hierarchical interconnect model and optimizes the clock topology generation. Several strategies and optimizations are implemented in different steps of the 3D CNS flow. Experiment results demonstrate the significant influence of inductance on 3D clock network, and the proposed approaches are proved to be effective.

## References

[1] Cadence. 2023. *Integrity 3D-IC Platform White Paper.* https://www.cadence.com/en_US/home/resources/white-papers/secured/system-driven-ppa-for-multi-chiplet-designs-wp.html
[2] Shivangi Chandrakar, Deepika Gupta, and Manoj Kumar Majumder. 2021. Role of through silicon via in 3D integration: Impact on delay and power. *Journal of Circuits, Systems and Computers (JCSC)* 30, 03 (2021), 2150051.
[3] Ting-Hai Chao, Yu-Chin Hsu, Jan-Ming Ho, and AB Kahng. 1992. Zero skew clock routing with minimum wirelength. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* 39, 11 (1992), 799–814.
[4] Walter James Condley, Xuchu Hu, and Matthew R Guthaus. 2010. Analysis of high-performance clock networks with RLC and transmission line effects. In *Proceedings of the 12th ACM/IEEE international workshop on system level interconnect prediction.* 51–58.
[5] Jason Cong and Cheng-Kok Koh. 1995. Minimum-cost bounded-skew clock routing. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, Vol. 1. IEEE, 215–218.
[6] Matthew R Guthaus, Gustavo Wilke, and Ricardo Reis. 2013. Revisiting automated physical synthesis of high-performance clock networks. *ACM Transactions on Design Automation of Electronic Systems (TODAES)* 18, 2 (2013), 1–27.

[7] Yehea I Ismail, Eby G Friedman, and Jose L Neves. 1999. Equivalent Elmore delay for RLC trees. In *Proceedings of Design Automation Conference (DAC).* 715–720.
[8] Guruprasad Katti, Michele Stucchi, Kristin De Meyer, and Wim Dehaene. 2010. Electrical modeling and characterization of through silicon via for three-dimensional ICs. *IEEE Transactions on Electron Devices (TED)* 57, 1 (2010), 256–262.
[9] Tak-Yung Kim and Taewhan Kim. 2011. Clock tree synthesis for TSV-based 3D IC designs. *ACM Transactions on Design Automation of Electronic Systems (TODAES)* 16, 4 (2011), 1–21.
[10] Tak-Yung Kim and Taewhan Kim. 2012. Resource allocation and design techniques of prebond testable 3-D clock tree. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)* 32, 1 (2012), 138–151.
[11] Kazuo Kondo, Morihiro Kada, and Kenji Takahashi. 2015. *Three-Dimensional Integration of Semiconductors: Processing, Materials, and Applications (Chapter 9).* Springer.
[12] Minghao Lin, Heming Sun, and Shinji Kimura. 2016. Power-efficient and slew-aware three dimensional gated clock tree synthesis. In *Proceedings of IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC).* 1–6.
[13] Tiantao Lu, Caleb Serafy, Zhiyuan Yang, Sandeep Kumar Samal, Sung Kyu Lim, and Ankur Srivastava. 2017. TSV-based 3-D ICs: design methods and tools. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)* 36, 10 (2017), 1593–1619.
[14] Tiantao Lu and Ankur Srivastava. 2015. Electromigration-aware clock tree synthesis for TSV-based 3D-ICs. In *Proceedings of Great Lakes Symposium on VLSI (GLVLSI).* 27–32.
[15] Tiantaoo Lu and Ankur Srivastava. 2017. Low-power clock tree synthesis for 3D-ICs. *ACM Transactions on Design Automation of Electronic Systems (TODAES)* 22, 3 (2017), 1–24.
[16] NCSU. 2011. FreePDK45. eda.ncsu.edu/freepdk/freepdk45.
[17] Deok Keun Oh, Mu Jun Choi, and Ju Ho Kim. 2019. Thermal-aware 3D symmetrical buffered clock tree synthesis. *ACM Transactions on Design Automation of Electronic Systems (TODAES)* 24, 3 (2019), 1–22.
[18] Xiaoning Qi, Gaofeng Wang, Zhiping Yu, R.W. Dutton, Tak Young, and N. Chang. 2000. On-chip inductance modeling and RLC extraction of VLSI interconnects for circuit simulation. In *Proceedings of the IEEE Custom Integrated Circuits Conference (CICC).* 487–490.
[19] P.J. Restle, T.G. McNamara, D.A. Webber, P.J. Camporese, K.F. Eng, et al. 2001. A clock distribution network for microprocessors. *IEEE Journal of Solid-State Circuits (JSSC)* 36, 5 (2001), 792–799.
[20] Yousef Safari and Boris Vaisband. 2023. A Robust integrated power delivery methodology for 3-D ICs. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems (TVLSI)* 31, 3 (2023), 287–295.
[21] Kamineni Sumanth Kumar and John Reuben. 2016. Minimal buffer insertion based low power clock tree synthesis for 3D integrated circuits. *Journal of Circuits, Systems and Computers (JCSC)* 25, 11 (2016), 1650142.
[22] Raja Swaminathan. 2021. *Advance Packaging: Enabling Moore's Law's Next Frontier through Heterogeneous Integration.* https://hc33.hotchips.org/assets/program/tutorials/2021%20Hot%20Chips%20AMD%20Advanced%20Packaging%20Swaminathan%20Final%20%2020210820.pdf
[23] C. N. Sze. 2010. ISPD 2010 high performance clock network synthesis contest: benchmark suite and results. In *Proceedings of the 19th International Symposium on Physical Design (ISPD).* 143.
[24] Cliff N Sze, Phillip Restle, Gi-Joon Nam, and Charles Alpert. 2009. ISPD2009 clock network synthesis contest. In *Proceedings of International Symposium on Physical design (ISPD).* 149–150.
[25] Aida Todri-Sanial and Chuan Seng Tan. 2017. *Physical Design for 3D Integrated Circuits (Chapter 7).* CRC Press.
[26] Chia-Chun Tsai, Jan-Ou Wu, Chien-Wen Kao, Trong-Yen Lee, and Rong-Shue Hsiao. 2006. Coupling aware RLC-based clock routings for crosstalk minimization. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS).*
[27] Wei Wang, Vasilis F Pavlidis, and Yuanqing Cheng. 2020. Zero-skew Clock Network Synthesis for Monolithic 3D ICs with Minimum Wirelength. In *Proceedings of the 2020 on Great Lakes Symposium on VLSI (GLSVLSI).* 399–404.
[28] Shyh-Chyi Wong, Gwo-Yann Lee, and Dye-Jyun Ma. 2000. Modeling of interconnect capacitance, delay, and crosstalk in VLSI. *IEEE Transactions on Semiconductor Manufacturing (TSM)* 13, 1 (2000), 108–111.
[29] John Wuu, Rahul Agarwal, Michael Ciraula, Carl Dietz, Brett Johnson, et al. 2022. 3D V-Cache: the implementation of a hybrid-Bonded 64MB stacked cache for a 7nm x86-64 CPU. In *Proceedings of IEEE International Solid-State Circuits Conference (ISSCC)*, Vol. 65. 428–429.
[30] Fan Yang, Minghao Lin, Heming Sun, and Shinji Kimura. 2017. Time-efficient and TSV-aware 3D gated clock tree synthesis based on self-tuning spectral clustering. In *Proceedings of IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS).* 1200–1203.
[31] Xin Zhao, Jacob Minz, and Sung Kyu Lim. 2011. Low-power and reliable clock network design for through-silicon via (TSV) based 3D ICs. *IEEE Transactions on Components, Packaging and Manufacturing Technology (TCPMT)* 1, 2 (2011), 247–259.