Cost-efficient 3D Integration to Hinder Reverse Engineering During and After Manufacturing

Peng Gu, Dylan Stow, Prashansa Mukim, Shuangchen Li and Yuan Xie
University of California Santa Barbara, CA, USA
{peng_gu, yuanxie}@ucsb.edu

Abstract—**Reverse engineering (RE) attacks pose a serious threat to the semiconductor supply chain. In this paper, we address this problem by proposing a design flow that leverages the unique capabilities of 3D integration to synergistically combine split fabrication and circuit camouflaging. First, 3D heterogeneous multi-chip integration is utilized to support efficient die-level split fabrication to protect against RE during manufacturing. Second, to prevent RE after manufacturing, a subset of the trusted foundry's die is obscured to significantly increase decamouflaging difficulty. A security-based cutsize-optimized partition algorithm is proposed to maximize the number of securely camouflaged gates on the trusted die while reducing the cutsize. Third, cost modeling is applied to demonstrate the cost effectiveness of the proposed 3D split fabrication flow compared with existing solutions. Across six widely used benchmarks, evaluations on 3D split fabrication designs between 15nm to 90nm processes show that the proposed design can effectively hinder practical RE attacks during manufacturing (Hamming Distance=30%) and after manufacturing (Complexity-to-Decamouflage=145), with minimum overheads to cutsize, footprint, and cost.**

I. Introduction

Reverse engineering (RE) is a key technique exploited by adversaries to carry out intellectual property (IP) piracy and hardware trojan injections either during or after manufacturing. Recently, several promising schemes have been proposed to counteract the threat of reverse engineering. Split fabrication [8] [9] [19] divides the circuit design into multiple parts to be fabricated across separate untrusted and trusted foundries such that an adversary in the untrusted foundry will find it computationally infeasible to derive the whole design in a reasonable time. Circuit camouflaging [14] selects gates from the original circuit to be obfuscated in layout such that the number of input patterns to decamouflage the entire circuit would increase exponentially. Another technique called logic locking [16] modifies an IC design by inserting key-gates such that it operates correctly only when a set of key inputs are set with the correct values.

Current limitations of these techniques are the inability to provide a holistic protection scheme against threats during and after manufacturing, as well as the extra overhead they induce. Split fabrication (front-end-of-line (FEOL) / back-end-of-line (BEOL) [9], or passive interposer [8] [19]) is ineffective after manufacturing because the adversaries can acquire the final chip and easily reverse engineer the metal-only layers of the trusted foundry partition. Furthermore, the technology gap between available trusted and untrusted processes forces either technology constraints or large area overheads during BEOL split fabrication. Circuit camouflaging [14] fails during manufacturing, as the untrusted foundry has the detailed layout information needed to fabricate the chip. The overheads and technology constraints of these two techniques are further explained in Sec. II. Logic locking [16] only partially protects against reverse engineering during manufacturing, since the untrusted foundry has access to the whole design and possesses the capability to reveal part of the design. It also introduces gates overhead as key-gates are inserted in the original netlist.

For the first time, we propose to use cost-efficient 3D integration (Sec. II-A) to combine the concepts of split fabrication (Sec. II-B) and circuit camouflaging (Sec. II-C) so that IP is secured against reverse engineering attacks during and after manufacturing. The unique opportunities provided by 3D integration can be leveraged to enhance security while reducing the overheads of existing split fabrication schemes. This scheme is cost-efficient, since it introduces minimal wire-length overhead where placement obfuscation is not needed. Cutsize overhead is also small compared with wire-based split fabrication since active devices can also be moved during partitioning. Our main contributions are as follows:

- We propose a design flow to combine circuit camouflaging and split fabrication that leverages cost-efficient 3D integration to effectively thwart reverse engineering during and after manufacturing (Sec. III and IV). The proposed technique lifts both transistors and wires onto the trusted die.
- An efficient security partitioning algorithm is designed to optimize security and cutsize (Sec. IV). Evaluation results show that it can effectively prevent proximity attacks from untrusted foundries during manufacturing, while significantly increasing the difficulty of circuit decamouflaging attacks after manufacturing, with minimal cutsize and area overheads (Sec. V).
- We quantify the cost effectiveness of the 3D integration used in our scheme against untrusted circuits and alternate security solutions. A structured ASIC reuse strategy is further explored for the trusted die to minimize the non-recurring mask cost overhead (Sec. VI).

This work was supported in part by NSF 1730309/1500848 and SRC/DARPA Center for Research on Intelligent Storage and Processing-in-memory (CRISP).

978-1-5386-7471-0/18/$31.00 ©2018 IEEE

II. Background and Related Work

A. 3D Integration

3D integrated circuit (3DIC) [5], in which multiple chips are fabricated independently and then assembled, has the benefits of high transistor density, shorter interconnect, and heterogeneous process integration. There are two types of 3DIC depending on the bonding style: (1) face-to-back (F2B) integration (Fig. 1(b)), where multiple dies are stacked in the same direction and connected by through-silicon-vias (TSVs) and micro-bumps (μbumps), (2) face-to-face (F2F) integration (Fig. 1(c)), where two dies are stacked face-to-face connected by only μbumps. Bonding may also be classified as wafer-to-wafer (W2W), in which wafers are bonded before dicing, or die-to-wafer (D2W), which enabled known good die testing before bonding. Another variant of 3DIC is passive 2.5D integrated circuit (2.5DIC) (Fig. 1(e)), in which multiple chips are placed side-by-side and interconnected by wires on a silicon interposer.

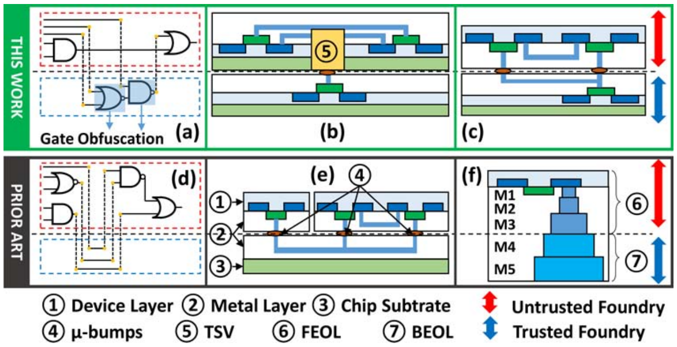


Figure 1. (a) 3D camo. split-fab, (b) face-to-back (F2B) 3D IC, (c) face-to-face (F2F) 3D IC, (d) wire lifting split-fab, (e) 2.5D passive interposer IC, (f) 2D IC

This paper assumes F2F W2W 3DIC (Fig. 1(c)) for split fabrication for the following reasons. Compared to F2B 3DIC, F2F 3DIC avoids TSV area overhead that occupies the area used for transistors. Compared with 2.5D passive interposer design, F2F 3DIC enables both transistors and wires to be lifted into the secure tier in trusted foundry. Furthermore this allows devices on the trusted tier to be camouflaged, thus adding protection after chips are shipped to market. W2W bonding is assumed due to the difficulty of validating partitioned circuits before bonding (an option with D2W) and because μbump pitches can be fabricated as small as $1\mu m$ [1].

B. Split Fabrication

Split fabrication [8] [9] [19] is mainly used to thwart reverse engineering attacks from a malicious foundry by limiting the adversary's access to only the partial design during outsourced IC production. Conventional split fabrication techniques are based on wire lifting as shown in Fig. 1(d). In 2D designs (Fig. 1(f)), the circuit is split into Front-End-Of-Line (FEOL) parts, which contain the transistors and associated low level metal layers (M1-3 in the example) fabricated in the untrusted foundry, and Back-End-Of-Line (BEOL) parts, which contains only high level metal layers (M4-5 in the example) manufactured in the trusted foundry. For passive interposer 2.5D designs, selected wires from the netlist are lifted and routed through the passive interposer produced in the trusted foundry, and the rest of the metal layers and the transistors are made in untrusted foundries.

Existing split fabrication techniques have three major limitations that can be addressed with the proposed 3DIC design flow. First, the hidden metal wires provide poor obfuscation protection after product shipping, since the malicious foundry has access to the whole product and can easily reverse engineer the metal layers. 3DIC can incorporate IC camouflaging on the trusted device layer (Fig. 1(a)) to thwart this attack. Second, since only metal wires are hidden, conventional approaches require placement and routing obfuscation [17] to prevent proximity attacks, which induces considerable wirelength overheads. With our method, the transistors and wires hidden in the trusted partition are a total black box and we carefully choose these black boxes to be interfered with each other. Even if the circuit topology is revealed, it is infeasible to decamouflage the circuit given the exponential combination of various possibilities for the black boxes. Third, the strict pitch and size matching required in BEOL / FEOL split design limits the trusted foundries' ability to use the untrusted foundries' advanced process with large technology gap. 3DIC enables heterogeneous integration of different technology nodes with μbumps.

C. IC Camouflaging

Circuit camouflaging [14] is a layout-level technique that hampers an attacker from reverse engineering after product shipping by employing dummy contacts into layout. Standard cells can be obfuscated using a mix of real and dummy contacts so that multiple gate types appear similar (in our case XOR, $NAND$ and NOR gate). This method alone is unsafe during manufacturing, since the untrusted foundry has full access to the original layout files and specific contact connectivity. Using 3DIC split fabrication, standard cell camouflaging can be implemented in the trusted foundry as shown in Figure 1 (a). How to select the gates to be camouflaged while reducing cutsize (Sec. IV) and the total footprint considering the camouflaging overhead (Sec. V) will be further addressed.

III. Security Assumption

A. Attack Model and Algorithm

During Manufacturing: We assume adversaries are able to view and modify the layout information (e.g. GDSII file) of the partial design outsourced to the untrusted foundries. However, the confidentiality, authenticity, and integrity of the circuit design are guaranteed, and no high-level information (e.g. netlist) can be leaked to adversaries. The final assembly, testing, and packaging of the chip are also implemented in the trusted foundries.

For proximity attacks [19], the adversaries in an untrusted foundry will first use wirelength optimized heuristics during placement and routing to find out the connections between wires. Since active devices in the trusted partition are unknown, they will assume the missing logic to be a black box for brute force logic profiling. The adversaries can use test patterns provided by the designers to match whether their assumed logic block box model is correct.

After Manufacturing: After acquiring the whole chip product, the adversaries can carry out depackaging, chip decoupling, wafer thinning, probing of internal access points, and netlist extraction to infer the hidden trusted design.

For circuit decamouflaging attacks [4] [14], the adversaries get two copies of the original design, one for input-output pattern generation and the other for reverse engineering. After delayering and imaging the product chip, the attackers are able to obtain the most netlist information, other than the functionality of the camouflaged gates on the trusted partition. Then, attackers can carry out a brute force attack [14] where each possible combination of the camouflaged gates is verified against correct input-output pairs, or an SAT attack [4] where a discriminative set is first calculated to reduce the input space and then applied on the circuit to eliminate all the impossible combinations of the camouflaged gates. Circuit testing techniques such as activation, sensitization, and automatic pattern generation tools [12] can assist this attack. In this work we assume a brute force attack to demonstrate the effectiveness of our proposed method.

B. Security Metrics

Hamming Distance (HD) [19] is a widely adopted metric to evaluate the protection against proximity attacks. Given the same input vector, HD equals the normalized number of different output bits between the original netlist and the reconstructed netlist from the partial circuit. A set of input vectors is used for evaluations between the function of the original netlist and the predicted function of the reconstructed netlist.

Complexity-to-Decamouflage (CtD), defined as the computational effort and the number of test patterns needed to learn the netlist using either brute force methods [14] or SAT based attacks [4], is assumed for quantifying the difficulty to decamouflage an obfuscated circuit. For an SAT attack, an SAT-hard clique is synthesized to increase the attacker's computational effort to determine the discriminative set, or a smart strategy for camouflaging is employed to increase the attacker's query complexity. Based on the attack model in Sec. III-A, we emphasize that CtD is an indication of attack hardness in log scale instead of an exact measurement of the number of steps for a certain attack. For both attacks, we assume that n gates have been securely camouflaged, each of which has m possibilities. We conservatively assume the computational complexity increases exponentially with the size of the camouflaged gate set ($O(log_{10}(m^n))$).

IV. Design Flow

Our proposed 3DIC based secure split fabrication design flow consists of four stages as shown in Fig. 2(a).

In the first stage, based on a security concept called gate interference (Sec. IV-B1), the largest interference graph will be selected to form a clique. Before partitioning, the designer needs to provide three parameters: (1) Partition Ratio ($pratio$), which is determined by the technology ratio used at trusted and untrusted die, (2) Security Requirement (N_secure_{min}), which is the minimum number of fully interfered camouflaged gates that are placed on the trusted die, and (3) Overhead Constraint ($CutSize_{max}$), which is the maximum partition cutsize allowed. The security optimized min-cutsize algorithm (Sec. IV-B3) will use the largest clique to initialize the partition and optimize security and cutsize under the above constraints. In the second stage, Partition 1 netlist ($C_{trudsted}$) will be synthesized according to Camo Cell List ($C_{camouflaged}$) and the gate camouflaging strategy adopted by the trusted foundry. In the third stage, if the timing and performance of the wirelength optimized placement and routing cannot be satisfied, then $pratio$ and N_secure_{min} will be relaxed in the first stage to re-generate the partition. This process will loop until a satisfying partition is achieved. Then the final split fabrication is carried out and assembly as well as testing will be done in the trusted foundry.

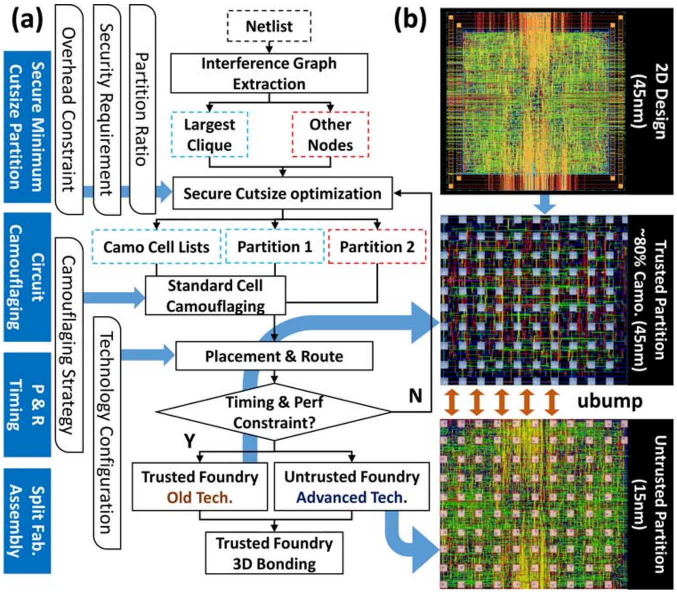

(a)
Secure Minimum Cutsize Partition
Overhead Constraint
Security Requirement
Partition Ratio
Netlist
Interference Graph Extraction
Largest Clique
Other Nodes
Secure Cutsize optimization
Circuit Camouflaging
Camouflaging Strategy
Camo Cell Lists
Partition 1
Partition 2
Standard Cell Camouflaging
P & R Timing
Technology Configuration
Placement & Route
Timing & Perf Constraint?
N
Y
Split Fab. Assembly
Trusted Foundry Old Tech.
Untrusted Foundry Advanced Tech.
Trusted Foundry 3D Bonding
(b)
2D Design (45nm)
Trusted Partition ~80% Camo. (45nm)
ubump
Untrusted Partition (15nm)

Figure 2. (a) Secure 3D split-fab design flow, (b) Our camouflaged 3D split-fab layout of a subset of processor [2]. Square array represents $\mu bumps$

A. Problem Formulation

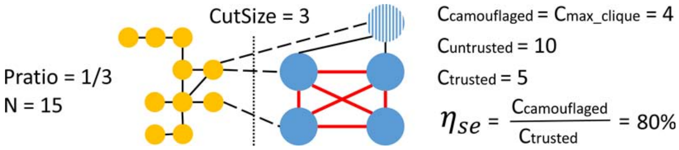

CutSize = 3
Pratio = 1/3
N = 15
Ccamouflaged = Cmax_clique = 4
Cuntrusted = 10
Ctrusted = 5
ηse = Ccamouflaged / Ctrusted = 80%

Figure 3. Illustration of maximizing interference graph while reducing cutsize during partition

Given the netlist of a circuit C with gate count N, partition ratio $pratio$, maximum cutsize $CutSize_{max}$ and minimum number of fully interfered gates N_secure_{min}, find partitions $C_{trusted}$, $C_{untrusted}$, and camouflaged gate list $C_{camouflaged}$, so that: (1) the largest interference graph C_{max_clique} on the trusted partition $C_{trusted}$ is maximized and size of C_{max_clique} is larger than N_secure_{min}, (2) the cutsize between the partition $C_{trusted}$ and $C_{untrusted}$ is minimized and that cutsize is smaller than $CutSize_{max}$. Selection Efficiency (η_{se}), which is the ratio of $C_{camouflaged}$ and $C_{trusted}$, is used to characterize the security strength of the partition method. Fig. 3 illustrates an example of the partition problem, where each gate is represented by a node and interference (Sec. IV-B1) is shown as an edge in the graph. Note that not every interference means a physical gate connection (wire).

B. *Secure Min-Cutsize Partition Algorithm*

1) Interference Graph: Gate interference [14] is utilized to increase the difficulty of circuit decamouflaging. If gate A is said to logically interfere with gate B, then either one of the following two conditions needs to be satisfied: (1) the inputs of A is on the output path of B, or if inputs of B is on the output path of A, (2) the primary output of A and B converges. In order to decamouflage a target gate, the attacker needs to feed several input patterns to the gate (activation) and observe the output of that gate from one or several primary output ports (sensitization). To maximally enhance the effectiveness of circuit camouflaging, the largest interference graph (theoretical maximum complexity) is extracted from the original netlist, where every gate in that graph is interfered with each other, and camouflaged using the strategy introduced in Sec. II-C.

2) Cutsize Optimized Algorithm: Fiduccia-Mattheyses-Sanchis (FMS) algorithm [3] is chosen to divide a given netlist into two partitions under a given ratio to minimize the cutsize. At the beginning, the netlist is converted to a hypergraph representation and two partitions are initialized according to the given partition ratio ($pratio$). A gain bucket (GB) is constructed to record the cutsize gain of moving a gate from one partition to the other. The partition process consists of multiple passes, during each of which the gate with the highest gain in the gain bucket is selected, moved and locked. We choose FMS as a baseline algorithm since its wide adoption on circuit partition task and its flexibility to control the gate movement under a given partition ratio.

Algorithm 1: Secure Min-Cutsize Algorithm

```
Input: $C$, $N$, $pratio$, $ratio_{off}$, $CutSize_{max}$, $N\_secure_{min}$
Output: $C_{trusted}$, $C_{untrusted}$, $C_{camouflaged}$
Data: $GB_1$, $GB_2$, $I$
$Init(C_{trusted}, I, GB_1)$, $Init(C_{untrusted}, \overline{I}, GB_2)$;
if $(size(I) > N \cdot pratio)$ then
    while $(size(GB_1) > 0)$ do
        Select gate $Gi$ of the highest gain from $GB1$;
        If move possible, update and lock;
else
    while $(size(GB_2) > 0$ do
        Select gate $Gi$ of the highest gain from $GB_2$;
        If move possible, update and lock;
Find max gain move seq. while $size(C_{trusted}) \geqslant N\_secure_{min}$;
Update $C_{trusted}, C_{untrusted}, GB_1, GB_2$;
if $(|\frac{size(C_{trusted})}{N_{trusted}} - 1| > ratio_{off} \ || \ cutsize \ > \ CutSize_{max})$ then
    Merge $GB_1$ and $GB_2$ to $GB$;
    Start FMS partition until $ratio_{off}$ and $cutsize$ is satisfied;
Extract largest $I$ from $C_{trusted} \rightarrow C_{camouflaged}$;
```

3) Proposed Algorithm: A secure min-cutsize partition algorithm is proposed to ensure both maximum security against decamouflaging attack (Sec. IV-B1) and minimum cutsize (Sec. IV-B2) as shown in Alg. 1. The algorithm has three stages:

①**Partition Initialization:** The largest interference graph I and its complement $\overline{I}$ is extracted from original netlist. The trusted (untrusted) partition and associated gain bucket is initialized using I ($\overline{I}$);

②**Unidirectional Gate Movement:** Depending on whether the size of I is larger than the number of gates on trusted partition or not, each time one gate of the highest move gain is selected from a gain bucket and locked. The unidirectional maximum gain move sequence is calculated under the constraint that the gates on trusted partition ($C_{untrusted}$) are more than the minimum required interference graph size (N_secure_{min}).

③**Bidirectional Gate Movement:** If the partition ratio is not satisfied, or the cutsize is larger than the limit, the two gain buckets (GB_1, GB_2) is merged and multiple cutsize optimization rounds (FMS) are allowed to exchange gates between these two partitions. A final satisfying result with the largest η_{se} is selected.

V. Evaluation

A. *Experiment Setup*

We demonstrate the effectiveness of the proposed 3DIC split fabrication flow against proximity attacks [13] during manufacturing and brute-force circuit decamouflaging attacks [14] after product shipping. Six circuits covering a wide spectrum of gate counts from ISCAS'85 [6] and ITC'99 [2] benchmark sets are evaluated. An open-source multi-pass partitioning software [3] is adapted to implement the proposed secure partitioning algorithm. The Hamming Distance used for proximity attacks is computed by Icarus Verilog [18] and the Complexity-to-Decamouflage is evaluated using an automatic test pattern generation tool [12] based on the PODEM algorithm.

B. *Security and Cutsize Improvement*

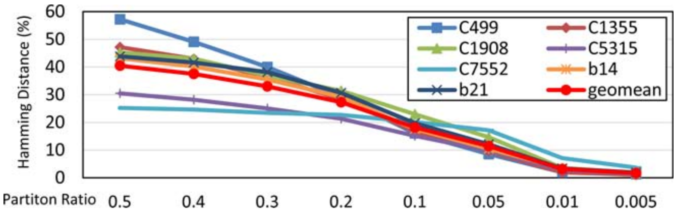


Figure 4. HD for different partition ratios

During manufacturing, Fig. 4 demonstrates the strength of our split fabrication method against proximity attacks. It is assumed that the missing circuit is replaced with a random logic black box after the attackers have figured out the connectivity of the untrusted partition using placement and routing heuristics. During the experiment, 100 random combinations of the black boxes are used and each combination is tested using 1000 test patterns. The final result is averaged across all the test patterns and combinations. It is shown from the figure that the geometric mean of HD across all benchmarks increases linearly with the trusted die partition ratio, and a reasonable split fabrication scheme between $32nm/15nm$ processes can achieve an average $HD = 28\%$ and an even split-fab ratio can have a very high average $HD = 41\%$.

Table I

CUTSIZE AND SECURITY EVALUATION OF PROPOSED ALGORITHM WITH DIFFERENT PARTITION RATIO

bench -mark	#gates	32nm/15nm, pratio=0.2				45nm/15nm, pratio=0.1				65nm/15nm, pratio=0.057				90nm/15nm, ratio=0.03				180nm, pratio=0.5 [19]	
		Cutsize		CtD		Cutsize		CtD		Cutsize		CtD		Cutsize		CtD		Cutsize	
		base.	**this**	base.	**this**	base.	**this**	base.	**this**	base.	**this**	base.	**this**	base.	**this**	base.	**this**	normal	secure
C499	202	12	14	30	33	9	12	18	24	9	14	13	14	5	5	7	7	16	45
C1355	546	7	12	48	50	4	5	23	25	5	5	13	15	2	4	8	8	16	43
C1908	880	15	17	67	71	10	11	32	38	6	6	12	22	4	7	3	12	35	37
C5315	2307	19	34	113	178	11	16	52	75	7	12	33	50	3	7	9	24	30	168
C7552	3512	15	43	191	201	15	37	75	151	9	15	47	98	9	10	58	74	25	155
b14	8567	75	87	63	576	46	87	26	426	47	93	18	241	27	72	20	124	99	386
b21	17482	99	101	65	569	80	102	48	569	53	78	21	529	29	76	23	82	-	-
geomean	1976	22.2	32.3	70.3	145.5	15.5	23.1	35.1	93.5	12.3	17.7	20.0	60.6	7.0	12.7	12.1	27.5	29.5	94.7

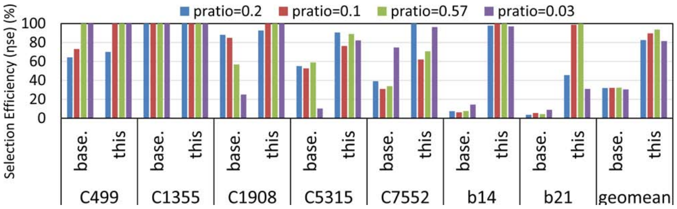


Figure 5. Selection Efficiency (η_{se}, defined in Sec. IV-A) comparison for different partition ratio

After manufacturing, Table I shows that our method can achieve high CtD against brute-force decamouflaging attacks with low cutsize overhead compared to the cutsize-optimized algorithm. When the circuit size is small ($< 1000\ gates$), the improvement of CtD is not significant ($2x$) but when the circuit size is very large ($\sim$ 10000 $gates$), the improvement of CtD can rise to $100x$. The main reason is that our method can achieve much higher Selection Efficiency (η_{se}) compared with the cutsize-optimization baseline as shown in Fig. 5. Given the same partition ratio, for large circuits our proposed method can put more securely camouflaged gates on the trusted partition ($80\% \sim 90\%$) than the baseline ($\sim 30\%$). Also, Table I demonstrates that our method incurs much lower cutsize overhead compared to the 2.5D passive interpose scheme [19], mainly due to the smaller partition ratio and the flexibility of moving both wires and gates during partitioning.

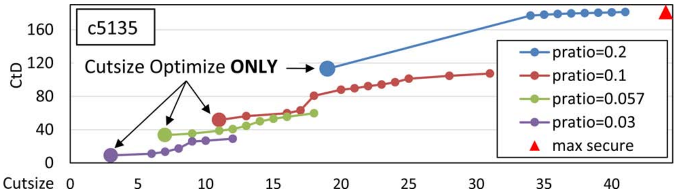
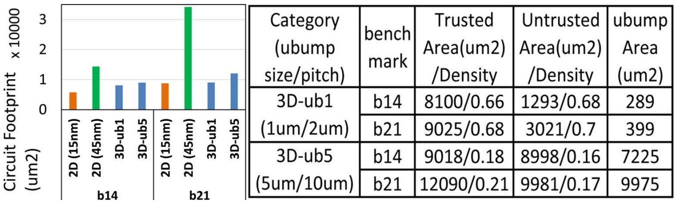


Figure 6. Design space for cutsize and security

We also demonstrate that our method allows for a flexible trade-off: allowing a small cutsize overhead for security improvement as shown in Fig. 6. Given a circuit (c5315 as an example), the maximum security level (red triangle) can be achieved by putting the largest interference graph on the trusted tier with large cutsize overhead however. To attain best cutsize, the cutsize-optimized design results in very low CtD under a fixed partition ratio. Alg. 1 allows changing both $CutSize_{max}$ and N_secure_{min} to flexibly explore the design space. For example, using 45nm/15nm ($pratio$=0.1) processes, by moving the point rightward, linear cutsize overhead (20) could be traded for exponential increase in decamouflaging complexity ($\Delta CtD = 56$).

C. Overhead Analysis



Category (ubump size/pitch)	bench mark	Trusted Area(um2) /Density	Untrusted Area(um2) /Density	ubump Area (um2)
3D-ub1 (1um/2um)	b14	8100/0.66	1293/0.68	289
	b21	9025/0.68	3021/0.7	399
3D-ub5 (5um/10um)	b14	9018/0.18	8998/0.16	7225
	b21	12090/0.21	9981/0.17	9975

Figure 7. Area analysis result

Area. We have 2 baselines for each of the 2 large circuit benchmarks, where the whole chip is fabricated in the trusted foundry ($2D(45nm)$) or untrusted foundry ($2D(15nm)$) as shown in Fig. 7. For our design, we assume 80% of the gates on the trusted die have been camouflaged with $4x$-area standard cell [14] and use two setups [1]: $1\mu m$ ubump with $2\mu m$ pitch ($3D$-$ub1$), and $5\mu m$ μbump with $10\mu m$ pitch ($3D$-$ub5$). For 3D split fabrication, the maximum of either trusted die area, untrusted die area, or μbump area is used as the circuit footprint. The final result shows on average our method has 22.6% area overhead compared with $2D(15nm)$ design (advanced tech., no trust, no camouflaged gates) and 52.7% area saving compared with $2D(45nm)$ design (old process, fully trusted, no camouflaged gates). The area overhead of our design is partially attributed to the less optimized Placement and Routing tool for 3D designs, and in future work we could improve the $P\&R$ and μbump assignment to reduce this overhead.

Performance. Compared to RE-resilient circuits manufactured in older trusted processes, our solution will likely realize performance gains from the adoption of faster process technologies. The inter-die μbumps introduce minimal delay, likely less than 1 ps based on similar TSV delay values [7]. If the two processes demand different voltage levels, level shifters are required on each low-to-high transition, introducing delay on the order of $70ps$ [11]. However, the stagnation of voltage scaling means that most processes since $90nm$ are able to run at compatible voltages [10] without level shifters. Full analysis of performance optimized partitioning across heterogeneous processes is for future work.

VI. Cost Effectiveness

BEOL split fabrication schemes are limited by the interconnect pitch differences of the untrusted and trusted processes. Because available trusted processes are several nodes older, most split fabrication solutions will introduce significant route-induced area overheads when using the latest technology.

A key benefit of the security solution proposed in this work is that the circuit can be primarily manufactured, without significant reduction in gate density, using the most advanced process technologies with the best performance, efficiency, and cost per transistor, even when the associated foundry is not trusted. However, our scheme requires the additional bonding of the trusted die produced in the trusted foundry's older process. Due to the heterogeneous processes used in this scheme, silicon area alone is an inaccurate proxy for cost, so it is necessary to directly model the costs of the proposed 3D system versus equivalent systems in untrusted advanced processes and in trusted older processes.

A. *Manufacturing Cost Effectiveness*

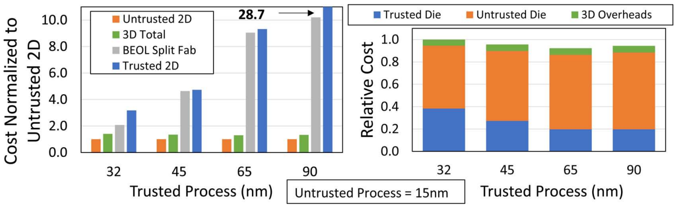


Figure 8. (a) Total cost comparison, (b) 3D split-fab cost decomposition

To determine the recurring manufacturing costs, a 3D cost model was developed following the methodology outlined by Stow et al. [15]. We conservatively assume wafer-to-wafer bonding without pre-bond validation, due to the complexity of validating the partitioned paths before bonding. The costs of the 2D dies are determined using the wafer costs [10], estimated die area from gate count, the number of dies per wafer, and the yield using process defect density. Next, 3D cost overheads are added to account for the extra required processing, including μbump deposition, through-silicon vias for substrate connectivity, and loss from bond yield. Relative costs are shown in Fig. 8(a) for: 1) untrusted advanced process only 2) our solution with 3D heterogeneous processes 3) BEOL split fabrication with area multiplier of M1/M4 pitch ratio squared 4) trusted process only.

As shown in Fig. 8 for a circuit size of 100 million gates, our solution introduces an average cost overhead of only 34% versus the untrusted circuit. The 3D solution is also significantly less expensive than the BEOL split fabrication or trusted process circuits, which both drastically increase area in older processes. Total cost is almost constant across trusted process selection, despite the changing constituent ratios shown in Fig. 8(b), with max variation less than 5% from the mean. Therefore, our 3D solution cost-effectively enables advanced untrusted processes even when their is a large technology gap with the trusted foundry's process.

B. *Mask Cost Overhead and Reduction*

The proposed solution also introduces a non-recurring cost overhead due to the additional set of masks for the trusted die. However, masks costs have risen with recent processes, so the trusted mask overhead is less than the advanced untrusted mask cost. This mask overhead for the trusted die can be almost completely removed by employing reusable methodologies, such as structured ASICs, that amortize most of their mask costs across a large volume. Although structured ASICs have a reduced gate density versus standard cell ASICs, the partition ratio in our security solution can be adjusted to account for this delta by moving more gates to the advanced process, thus minimizing mask overhead with minimal impact on recurring cost and security.

VII. Conclusion

Existing techniques to counter IC reverse engineering attacks can only provide partial protection during and after manufacturing, and they introduce technology constraints or various sources of overhead. Utilizing 3D integration's capability for heterogeneous technology integration and die stacking, we propose to securely select a partition to be fabricated in the advanced but untrusted foundry, while camouflaging part of the circuit at the trusted foundry to provide protection after manufacturing. Evaluation results show that our method can effectively improve security and optimize the cutsize with small overheads. Further, 3D cost analysis verifies that our method is cost-efficient compared to prior solutions.

References

[1] E. Beyne. The 3-d interconnect technology landscape. *IEEE Design & Test*, 2016.

[2] F. Corno et al. Rt-level itc'99 benchmarks and first atpg results. *IEEE Design & Test of computers*, 2000.

[3] A. Dasdan et al. Two novel multiway circuit partitioning algorithms using relaxed locking. *TCAD*, 1997.

[4] M. El Massad et al. Ic decamouflaging: Reverse engineering camouflaged ics within minutes. In *NDSS*, 2015.

[5] P. Gu et al. Leveraging 3d technologies for hardware security: Opportunities and challenges. In *GVLSI*, 2016.

[6] M. C. Hansen et al. Unveiling the iscas-85 benchmarks: A case study in reverse engineering. *IEEE Design & Test of Computers*, 1999.

[7] H. Homayoun et al. Dynamically heterogeneous cores through 3d resource pooling. In *HPCA*, 2012.

[8] F. Imeson et al. Securing computer hardware using 3d integrated circuit (ic) technology and split manufacturing for obfuscation. In *USENIX Security*, 2013.

[9] M. Jagasivamani et al. Split-fabrication obfuscation: Metrics and techniques. In *HOST*, 2014.

[10] M. Khazraee et al. Moonwalk: Nre optimization in asic clouds. *SIGPLAN*, 2017.

[11] W. Liu et al. Enhanced level shifter for multi-voltage operation. In *ISCAS*, 2015.

[12] H.-K. Ma et al. Test generation for sequential circuits. *TCAD*, 1988.

[13] J. Magaña et al. Are proximity attacks a threat to the security of split manufacturing of integrated circuits? In *ICCAD*, 2016.

[14] J. Rajendran et al. Security analysis of integrated circuit camouflaging. In *CCS*, 2013.

[15] D. Stow et al. Cost and thermal analysis of high-performance 2.5d and 3d integrated circuit design space. In *ISVLSI*, 2016.

[16] P. Subramanyan et al. Evaluating the security of logic encryption algorithms. In *HOST*, 2015.

[17] Y. Wang et al. Routing perturbation for enhanced security in split manufacturing. In *ASPDAC*, 2017.

[18] S. Williams et al. Icarus verilog: open-source verilog more than a year later. *Linux Journal*, 2002.

[19] Y. Xie et al. Security-aware design flow for 2.5 d ic technology. In *TrustED*, 2015.