

Assessing Benefits of a Buried Interconnect Layer in Digital Designs

Liheng Zhu, Yasmine Badr, Shaodi Wang, Subramanian Iyer and Puneet Gupta

EE Department, University of California, Los Angeles

ABSTRACT—In sub-15nm technology nodes, local metal layers have witnessed extremely high congestion leading to pin-access-limited designs, and hence affecting the chip area and related performance. In this work we assess the benefits of adding a buried interconnect layer below the device layers for the purpose of reducing cell area, improving pin access and reducing chip area. After adding the buried layer to a projected 7nm standard cell library, results show ~9-13% chip area reduction and 126% pin access improvement. This shows that buried interconnect, as an integration primitive, is very promising as an alternative method to density scaling.

Keywords—Buried layer, interconnect, standard cell, device structure, 7nm node

I. INTRODUCTION

With feature dimensions reaching the nanometer scale, local metal layers have become extremely valuable routing resources since they are heavily used for standard cell routing and pin access. In FinFET technologies [1]-[3], the introduction of the Local Interconnect (LI), which is used to connect fins or gates to make a multi-fin or multi-poly device [4], helped reduce congestion on these layers. However, it is unlikely that adding more LI layers will give significant benefits since the contact (V_0) holes that are connected to the device layers are also highly congested. Moreover, pin access has also become one of the biggest challenges to scaling density, since technology scaling and the design of cells with compact area has made it extremely difficult for the routers to access the pins [5]. Metal design rules, have become limited by the lithography resolution. Thus, in a lot of cases, chip area is routing-limited, which reduces the potential benefit of technology scaling.

The use of more metal layers above the device layers for intra-cell routing adds to the severity of the congestion problem. This is because for every route on a higher metal layer, vias and landing pads have to be placed on the lower layer, leading to what is known as via blockage problem. Moreover, the landing pads must satisfy the minimum area rule, and thus routing resources on the lower layers are wasted. This problem is only getting worse as metal minimum area rules have not scaled as much as pitch rules (the latter being aided by multiple patterning). For example, Figure 1 shows three metal layers (M_1 - M_3), along with the required vias. Because of the congestion on M_1 and M_2 , the route was resumed on M_3 . However, landing pads are still needed on M_2 , consuming M_2 space. Earlier research [6] has studied the problem of via blockage, and it has been postulated that via blockage limits the benefit of increasing the number of metal layers. Moreover, it has been shown in [7] that the via blockage factor can be as high as 50% on the first metal layer.

In order to solve these problems, we propose using a buried metal layer in the standard cells. This buried metal layer

(M_{-1}) and its contact layer (V_{-1}), lie underneath the device layers. Figure 2(a) shows a cross-section of the traditional interconnect stack on top of the device layers, and Figure 2(b) shows the interconnect stack after adding the proposed buried layer under the device layers.

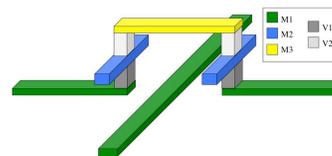


Figure 1. Via Blockage with the use of three metal layers. Notice the wasted area on M_2 due to the need for landing pads satisfying minimum area rule.

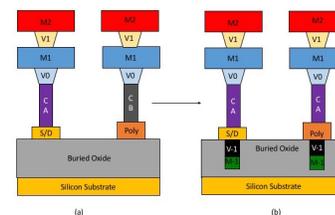


Figure 2. (a) Original stack of interconnect layers on top of device layers in SOI process (b) New stack with M_{-1} and V_{-1}

Having an interconnect layer under the transistors in the standard cells is expected to be much more beneficial than adding a new LI layer on top of the device layers or using the upper metal layers (M_2 and possibly higher) for intra-cell routing because of several reasons. First, as mentioned above, the V_0 layer, which is the main bridge from the device layers to the metal layers, has become congested as well. Second, using the higher metal layers will add to the via blockage problem, due to the need for vias and landing pads as explained earlier. Third, sharing a layer (e.g. M_2) for intra-cell and inter-cell routing makes pin access even more difficult for the router by blocking some of the tracks that would have otherwise been available for pin access. Fourth, using M_1 intensely for intra-cell routing has urged designers to create short pins, which again complicates pin access [5].

The idea of using a buried metal layer under silicon is not totally novel, but using it in logic by introducing it to the standard cell library to make transistor connections has not been studied before, to the best of our knowledge. A buried layer has been used in DRAM for buried word lines [8]. For many years, extensive work has been done in direct wafer bonding [9] to include a buried metal layer in the dielectrically isolated substrate [10]. It has been proven that the buried metal layer can provide potential benefits to SOI substrates in MOS IC's [11],[12]. The electrical performance of this buried layer has also been experimentally studied to verify that the silicon substrates including the buried metal layer can exhibit good device behavior [13]. Finally, a buried source/drain contact is proposed in [14] for FinFET devices. All these works have different contexts and purposes of

using a buried layer and their proposed fabrication methods can not be applied when the buried layer is used to make local connections in standard cells.

In this paper, we study the performance and density implications of a buried interconnect layer for random digital logic. The contributions of this work can be summarized as follows:

- This is the first work to propose and evaluate the buried interconnect layer concept for random digital logic. An algorithm is implemented to modify a standard cell library, in order to introduce the buried layer and reduce the area of the standard cells as a result.
- Pin-access benefits of the buried layer are evaluated. In addition, several benchmarks are synthesized, placed and routed using the standard cell library with the buried layer, in order to assess the chip area savings.
- TCAD and Spice simulations are used to evaluate the performance impact of the buried layer. Effect of the buried layer on the chip-level performance is also evaluated.

The rest of the paper is organized as follows: Section II describes the introduction of the buried metal layer into a 7nm standard cell library, followed by the cell-area analysis. Section III presents the pin access improvement and the chip-area analysis. Section 3 describes the process flow that we propose to manufacture the buried interconnect. Performance evaluation is shown in Section V. Finally, conclusions and future work are shown in Section VI.

II. BURIED LAYER IN STANDARD CELLS

How Buried Interconnect contacts the Transistors. All the work in this paper assumed SOI-based FinFET process¹. The buried interconnect layer and the buried contact layer lie underneath the device layers. As shown in Figure 2, the buried V_I connects the buried M_I layer to the gate and source/drain by going through the SOI buried oxide.

The buried interconnect connects to source/drain region through a V_I hole which goes through the buried oxide and contacts the source/drain region. A cartoon diagram of source/drain contacts through V_I is shown in Figure 3. With FinFET devices, the source/drain region is usually formed by merging fins by epitaxial growth of SiGe [15]. Therefore, the buried vias do not need to contact the thin fins directly (which would have been very hard to control due to the overlay error).

To contact the gates, the contact V_I is placed between two adjacent fins as shown in Figure 4 to reach the gate. This requires the contact width (c_w) to satisfy this constraint: $c_w < f_s - 2o_t$ where f_s is the fin to fin spacing and o_t is the thickness of the high-K oxide that is underneath the gate.

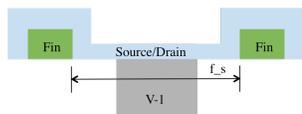


Figure 3. Contacting Source/drain contacts (V_I)

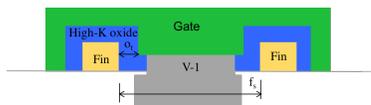


Figure 4. Contacting Gate contacts (V_I)

¹ The cost comparison of SOI vs. Bulk is outside the scope of this study. This work presents the idea on SOI. However, our future work will address the buried layer concept in a bulk Si process.

Experimental Setup. We introduced the buried layer to a projected 7nm FinFET standard cell library, from a leading IP provider. The cells have been modified in order to use a buried layer, as shown in the following sections. All M_I segments are horizontal, keeping the layer unidirectional.

Layout Changes. The buried layer (M_I) is used to completely replace the horizontal LI layer (CB), to make gate-to-gate connections. For example, Figure 5 shows the interconnect layers in a snippet of a hypothetical standard cell, where a gate-to-gate connection labeled x got transferred from CB to the buried layer M_I . Only one LI layer (CA) is used in the final standard cell library; CA is still preserved in order to connect the fins, create the power straps and connect the transistors to M_I . Therefore, CA is used for both gate and source/drain contacts like the contact layer in pre-LI technologies [16]. V_0 layer is used to connect M_I to the CA layer, as shown in Figure 2.

In this 7nm standard library, all of the I/O pins are on M_I layer, and the pins are accessed by dropping a via from M_2 to M_I . Thus, moving all the intra-cell M_2 routing segments to M_I is a top priority, because it is expected to improve pin access, and accordingly can result in chip-area reduction.

An algorithm is implemented to automate the relocation of M_I and M_2 routes to M_I . Nets containing more than two endpoints are broken down into multiple connections; each connection is between two endpoints. The resulting connections are handled as follows:

Gate-to-Gate Connection or Source/Drain-to-Gate Connection: An available track on M_I is picked for the route resulting in a horizontal M_I segment. Figure 5 shows a gate-to-gate route labeled z , which used to span M_I and M_2 , but got relocated to M_I (the input pin is kept on M_I , though).

Source/Drain-to-Source/Drain Connection: If both endpoints belong to P-FET transistors or both belong to N-FET transistors, then an available track on M_I is used for this connection. However, if one endpoint is a P-FET and the other is an N-FET, a vertical segment is needed which cannot be done on M_I , and thus part of the connection is done on M_I . For example in Figure 5, net y connects the drain regions of one P-FET and two N-FET transistors and it used to exist on layer M_I , thus the P-FET to P-FET connection got relocated to M_I but the P-FET to N-FET connection remained on layer M_I .

If relocating the route to M_I results in design rule violations or no available track is found on M_I , the route is kept on its original layer. In this library, all the routes on M_2 were successfully relocated to M_I , thus M_2 is no longer used for intra-cell routing. In addition some of the M_I routes were successfully moved to M_I , reducing the M_I congestion, as shown in Figure 5. However, some routes remain on M_I because they are I/O pins, vertical connections (M_I is horizontal layer), or because of the capacity of M_I .

Cell-level area reduction. Due to the relocation of M_I and M_2 to M_I as described in the previous sections, the consumed routing resources on M_I , M_2 and V_I have been greatly reduced. Therefore, standard cells which used to be routing-limited, meaning that their areas were determined by the need for more routing resources rather than by the transistors, are no longer routing-limited and their areas are reduced. This has been performed by removing the dummy polysilicon shapes (shapes on the polysilicon layer, that

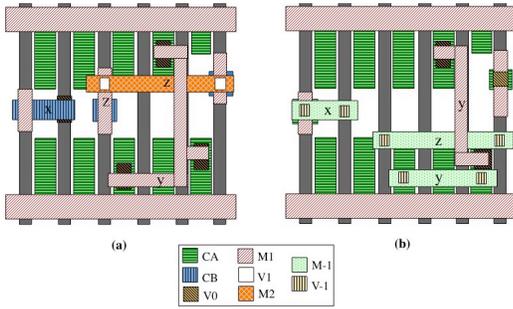


Figure 5. (a) Interconnect layers of a snippet of a hypothetical Standard Cell without buried layer (b) Same snippet with buried layer

are not gates) that used to exist to provide more space for routing. The cell-level area reduction algorithm attempts to delete the area occupied by such dummy poly shapes, and compacts the cells by shifting the contents of the cell. The routes that used to cross that area are then re-wired by being moved to other tracks, if they have design rule violations with the other routing segments. This is done as follows: M_1 routes with vertical steiner-tree trunks are changed by relocating the trunk to an available vertical track on M_1 . Horizontal connections passing through the eliminated area are made shorter easily without introducing problems.

Since the cell-level area reduction only targets cells which are routing limited, only the complex cells in the library can benefit from it. Accordingly, other cells whose areas are defined by the transistors do not get area reduction by using the buried interconnect. Thus in our experiments, which were performed on a 59-cell projected 7nm library, the three most complex cells witnessed an area reduction between ~6%-13%, as shown in Table 1. For the entire 7nm standard library, the average area reduction is around 0.48% over all cells since this is a very small library (59 cells) and most of the cells are simple ones that are not routing-limited.

Table 1. Results of the Standard Cell-level Area Reduction

Cell	Original area (μ^2)	New area (μ^2)	Area difference
LAT_X1	0.2916	0.2527	13.34%
MX2_X1	0.2138	0.1944	9.09%
SDFX_X1	0.62208	0.5832	6.25%

The cell area reduction could not have been done on the original library, without the buried layer, even if M_3 is used for intra-cell routing. Routing the reduced-area flip flop cell with three metal layers above the device (M_1 - M_3) and one local interconnect layer is impossible while routing it with one metal layer above device (M_1), one local interconnect layer (CA) and one buried layer is feasible. This supports our claim that the buried layer can solve problems and achieve benefits that above-device routing layers can not.

III. PIN ACCESS AND CHIP-LEVEL BENEFITS OF THE BURIED LAYER

In this section, we discuss two benefits gained from the addition of the new metal layer M_L : pin-access improvement and chip-level area reduction

A. Pin Access improvement. To quantify pin access, we use the metric proposed in [17], where a hit point is defined as the overlap of M_2 routing track and the I/O pin; and a valid hit point combination (VHC) is the set of hit points containing one hit point for each of the input and output pins, with no design rule violations. The total number of VHC is the used metric. The number of VHC in the original cells is constrained by intra-cell routes on M_2 . Since all the M_2 routes were successfully replaced

by M_L , the different ways to access a specific pin from both directions increase significantly. The hit point calculation algorithm has been applied to recursively count the total number of VHC for the cell libraries before and after using the buried layer. Since all pin access is assumed to be done through M_2 , only the cells that used to have M_2 routes can observe an improvement in pin access, when the buried layer is introduced. The results of the pin access improvement are summarized in Table 2. More M_2 used in the original standard cell results in greater pin access improvement that can be obtained due to M_L . The average pin access improvement is 126.32% for the listed four cells only.²

Table 2. Pin Access Improvement Results using Number of Valid Hit Point Combinations (VHC) as a metric

Cell	No. of VHC (Without Buried Layer)	No. of VHC (With Buried Layer)	Improvement (%)	No. of M_2 routes
SDFX_X1	35	84	140.0%	4
XNOR2_X1	2	8	300.0%	3
LAT_X1	30	46	53.3%	2
MX2_X1	183	205	12.02%	1

B. Chip-Area Saving. In order to check if these cell-level improvements can lead to final chip-level benefits, chip-level experiments have been performed. Using the new cell library made of 59 cells with M_L layer, three benchmarks were synthesized, placed and routed. A Layout Exchange Format (LEF) file reflecting the modified cells geometry, due to the buried layer, is used. The used test cases are FPU and MIPS from Open Cores as well as a Cortex M0 processor. These have been placed and routed using Cadence Encounter. The number of gates in each benchmark and the number of metal layers used in routing is shown in Table 3.

Table 3. Benchmarks used in Place and Route Experiments

Testcase	No. of Gates	No. of Routing Layers
Cortex M0	9800	4 ($M_2 - M_3$)
FPU	27140	3 ($M_2 - M_3$)
MIPS	7967	2 ($M_2 - M_3$)

The highest utilization factor, at which the design is routable with no design rule violations, is used in each experiment.

Table 4 shows the chip area reduction due to the decrease in congestion on M_1 and M_2 without decreasing the cell area, and the final area reduction due to the reduction in cell area as well as congestion relief on M_1 and M_2 . The final chip-level area saving is around 9 %-13%, due to the M_L buried layer.

Table 4. Results of the final chip area reduction

	Cortex M0	FPU	MIPS
Replacement of M_1 and M_2 without cell-level area reduction	9.9%	11.9%	7.5 %
Replacement of M_1 and M_2 with cell-level area reduction	11.8%	12.9%	8.9%

IV. MANUFACTURING PROCESS FLOW AND MP DECOMPOSITION

Process Flow. In order to pattern the buried interconnect and via layers, we propose the following process. Cross-sections of the proposed process steps are shown in Figure 6.

² The other cells in the library did not have any routes on M_2 . Reducing congestion on M_1 does not result in a change in the used pin access metric. Thus among the whole library, the average pin access improvement is 8.6% only.

are not analyzed because V_{L1} and M_{L1} layers are relatively sparsely utilized indicating that they have less coupling capacitance than M_1 , M_2 and V_0 .

The coupling effect may also introduce threshold voltage shift and leakage current increase of the fins that are electrically disconnected from the near vias. The introduced leakage decreases with the increase of V_{L1} -to-fin distance. A severe leakage increase only occurs when a via is placed very close to an active FinFET, e.g., a V_{L1} is placed between fins with 4nm distance can increase leakage by 20%. However, this never happens in our library, and the smallest V_{L1} -to-fin distance is greater than 40nm, which may have less coupling effect than other metal and via layers, e.g., some V_0 close to fins. Thus, the leakage penalty is negligible and we ignore it in the timing evaluation.

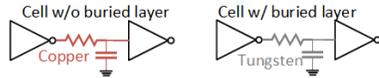


Figure 9. Simplified logic gate stage for standard cell delay change estimation due to using buried layer.

The buried layer is assumed to be made of Tungsten to be able to withstand the high temperature involved in wafer bonding. Thus, the buried layer has higher resistivity (5.6×10^{-6} ohm cm) than the copper metal layers (1.7×10^{-6} ohm cm). We estimate the effect of that on the propagation delay of the standard cells and chip-level performance. TCAD simulations on entire library will take infinite time and is out of the scope of this paper. We use a simplified logic gate stage (Figure 9) as in [22], [23], which contains a driving gate (e.g., inverter), a wire load (resistance and capacitance), and a gate load (e.g., inverter), to simulate the propagation delay change of standard cells after using the buried layer. The accuracy of the simplified gate model for chip-level speed estimation has been verified in [22], [23] against synthesis, placement, and routing. In our library, when a cell is redesigned with the additional buried layer, total copper wire length (M_1 and M_2) is reduced and the use of Tungsten (M_{L1} and CA) increases due to the routes that get relocated to M_{L1} from M_1 and M_2 . The reduced copper wire and two copper vias (V_0) are used as the wire load for simulating delay of standard cells without buried layer, while the increased tungsten wire and two tungsten vias (V_{L1}) serve as the wire load for cells with buried layer. The unit capacitance of copper wire is assumed to be same as tungsten, while in reality M_1 and M_2 layers are more utilized than M_{L1} and should have larger coupling capacitance than M_{L1} , indicating that the performance evaluation is pessimistic for the buried layer. The driving and load gate sizes vary from 3 to 12 fins according to standard cell size. The SPICE simulation results show that 23 out of 60 cells have decreased propagation delay after using the buried layer, while 39 cells see delay increase. The cell delay change ranges from -3.6% to 2.1%. In addition, on the average the buried layer led to 3.5% overall wire length reduction per cell, which explains the delay reduction of some standard cells.

To perform chip-level performance evaluation, a timing library is generated by applying the change in propagation delay to the original timing library. Timing analysis is performed after Place and Route using the modified library. The delay of the most critical path changes by 0.13%, 0%, and -0.01% in Cortex M0, MIPS and FPU respectively. Thus the chip-level performance change is too small and negligible.

VI. CONCLUSION & FUTURE WORK

A buried interconnect layer has been introduced to a standard cell library in order to alleviate the congestion on the traditional

interconnect layers. It has been shown that the buried layer can improve pin access by 126%, and save chip area by ~9-13%.

In this paper, cell-level area reduction is achieved mainly through removing the dummy polys in the standard cells, but the source/drain regions remain untouched. However, in our future work, we will add the flexibility of re-arranging the transistors or laying the transistors out from scratch, in order to make better use of the buried layer in terms of saving the cell-area. Our preliminary results of the flip flop cell using manual cell-design, shows that a further reduction of 3.1% of the flip flop cell area can be achieved if the buried layer was taken into consideration in the standard cell design phase. Moreover, intra-cell routes remaining on M_1 can be redesigned in order to be DP-compliant, thus leading to the use of DP to pattern M_1 instead of TP and saving one mask. In addition, we will investigate a second process flow where the buried interconnect layer is made of doped silicon and we will evaluate the effect of such a process on the performance. Finally, we will look into methods to integrate a buried interconnect layer into a bulk Si process.

REFERENCES

- [1] C. Auth, "22-nm fully-depleted tri-gate CMOS transistors," in Proc. IEEE Custom Integr. Circuits Conf. (CICC), 2012.
- [2] T. B. Hook, "Fully depleted devices for designers: FDSOI and finfets," in Proc. IEEE Custom Integr. Circuits Conf., 2012, paper 18-
- [3] C-H. Lin et al., "High performance 14 nm SOI FinFET CMOS technology with embedded DRAM and 15 levels of Cu metallization," in Proc. Int. Electron Devices Meet. (IEDM), 2014, pp. 74–76.
- [4] N. Lu, P. Kotecha, R. Wachnik, "Modeling of resistance in FinFET local interconnect". CICC 2014: 1-4
- [5] T. Taghavi, C. Alpert, A. Huber, Z. Li, G. Nam, S. Ramji. "New Placement Prediction and Mitigation Techniques for Local Routing Congestion", ICCAD'2010
- [6] Sai-Halasz, George. "Performance trends in high-end processors." Proceedings of the IEEE 83.1 (1995): 20-36.
- [7] Q. Chen, et al. "A compact physical via blockage model." Very Large Scale Integration (VLSI) Systems, IEEE Transactions on 8.6 (2000): 689-692.
- [8] Micron Technology, Method of making memory cell with vertical transistor and buried word and body lines, US Patent US5909618, 1997
- [9] W. P. Maszara, G. Goetz, A. Caviglia, and J. B. McKitterick, "Bonding of silicon wafers for silicon-on-insulator," J. Appl. Phys., vol. 64, pp. 4943–4950, Nov. 1988
- [10] J. B. Lasky, "Wafer bonding for silicon-on-insulator technologies," Appl. Phys. Lett., vol. 48, pp. 78–80, Jan. 1986
- [11] Y. Arimoto, H. Horie, N. Higaki, M. Kojima, F. Sugimoto, and T. Ito, "Advanced metal oxide semiconductor and bipolar devices on bonded silicon-on-insulator," J. Electrochem. Soc., vol. 140, pp. 1138–1143, Apr. 1993.
- [12] E. D. Nowak, L. Ding, Y. T. Loh, and C. Hu, "Speed, power, and yield comparison of thin bonded SOI versus bulk CMOS technologies," IEEE Int. SOI Conf., 1994.
- [13] W. L. Goh, J. H. Montgomery, S. H. Raza, B. M. Armstrong, "Electrical Characterization of Dielectrically Isolated Silicon Substrates Containing Buried Metallic Layers", IEEE Electron Device Letters, Vol. 18, No. 5, May 1997
- [14] H. Zang, C. Yeh, T. Yamashita, and V. Basker, V., Globalfoundries Inc. and International Business Machines Corporations, 2013. "Buried local interconnect in finfet structure." U.S. Patent Application 14/135,716.
- [15] H. Kawasaki, V. S. Basker, T. Yamashita, C. H. Lin, Y. Zhu, J. Faltermeier, S. Schmitz et al. "Challenges and solutions of FinFET integration in an SRAM cell and a logic circuit for 22 nm node and beyond." IEDM 2009
- [16] D. Jansen. "The electronic design automation handbook". Springer, 2010.
- [17] Xiaoqing Xu, Bei Yu, Jih-Rong Gao, Che-Lun Hsu, and David Z. Pan, "PARR: Pin Access Planning and Regular Routing for Self-Aligned Double Patterning", ACM/IEEE Design Automation Conference (DAC), 2015
- [18] A. Gharbi, R. Tiron, M. Argoud et al., "Contact holes patterning by directed self-assembly of block copolymers: what would be the bossing plot", Proc. SPIE 9049, Alternative Lithographic Technologies VI, 90491N, 2014
- [19] R. S. Ghaida and P. Gupta, "Role of design in multiple patterning: technology development, design enablement and process control," DATE, 2013
- [20] J. Schoota, K. Schenau, G. Bottigliera, K. Troosta, J. Zimmermann, S. Migurac, B. Kneerc, J. Neumann, W. Kaiser, "EUV High-NA scanner and mask optimization for sub 8 nm resolution", SPIE Photomask Technology, 2015
- [21] Marie Kryszak*, Michael Leeson, Eungnak Han, et al., "Extending resolution limits of EUV resist materials", SPIE Lithography symposium, 2015
- [22] Wang, Shaodi, et al. "PROCEED: A Pareto optimization-based circuit-level evaluator for emerging devices." ASPDAC. IEEE, 2014.
- [23] Shaodi Wang, A. Pan, Chi On Chui and P. Gupta, "PROCEED: A Pareto Optimization-Based Circuit-Level Evaluator for Emerging Devices," in IEEE TCAD, vol. 24, no. 1, 2016.
- [24] L. Liebmann, A. Chu, and P. Gutwin. "The daunting complexity of scaling to 7nm without EUV: Pushing DTCO to the extreme." In SPIE Advanced Lithography, 2015
- [25] P. A. Kearney, O. Wood, E. Hendrickx, et al. "Driving the industry towards a consensus on high numerical aperture (high-NA) extreme ultraviolet (EUV)". In SPIE Proceedings Extreme Ultraviolet (EUV) Lithography V, volume 9048, 2014

