nature portfolio

Peer Review File

Monolithic three-dimensional integration of silicon transistors

Corresponding Author: Professor Qing Cao

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
A. the approach of monolithic three-dimensional integration of complementary single-crystalline silicon layers is new and original.

B. A couple of weak points in this work:
(i) The main issue is, the authors have stated that the method of layer transfer introduced here would lead to higher integration density of devices. However, in the device section they have shown that one device layer (tier) has only 625 transistors on an area of 40 x 40 mm2, which is extremely low.
One can only speculate what the reason is for that low integration density but I suspect the release holes cause the problems. You have to put the transistor in an area far away from the release hole? But even if circuit designers would be able to use a design kit for those transistors I cannot see how to deal with the holes. Is the size of a circuit limited to 85 x 85 µm2? In my opinion it is too small for VLSI.
(ii) Another disadvantage is the preparation of the silicon layers for the transfer. You have to sacrifice one SOI wafer for each Si sheet you want to transfer. When stackin it as suggested with many tiers the cost and the effort will be enormous.

C. To perform the “uniform” doping by diffusion, high consumption of energy is required. Somewhere in the text, the authors claim that the doping is uniform. However, later on it is stated that the variations in threshold voltage and sub threshold swing might be caused in doping inhomogeneity in the layers. An appropriate testing the doping homogeneity would be a four point probe measurement across the wafer. A mapping would answer the question if it is homogenous or not.

D. The authors show statistics on their devices. The device variability across the wafers is too big.
Another point not discussed, is that for lower tiers, the threshold voltage seems to change and also the variation across the wafer is increasing.

E. In the conclusion the authors state that their process shows good uniformity and performance like FEoL silicon MOSFETs. In FEoL MOSFETs the variation of threshold voltage is in the range of couple mV and the variation in subthreshold slope is in the range of 1-2 % across a whole 200 or 300 mm wafer.
The 3D logic circuits just consists of a small number of transistors and only represent the basic circuits. To evaluate such a technique it should be appropriate to show circuits with thousands of transistor to also judge on the yield.

F. Information about the yield would be required. How is the yield on a wafer in the different tiers? How is the yield on whole wafer and also from lot to lot.

G. The references are adequate except the fact that the authors listed the works which dealing low dimensional nanomaterials just by naming them 10 – 22 in the abstract. But later in the table in Fig. 9 the papers are discussed in detail.

H. There are too many papers cited in the abstract.

Referee #2

(Remarks to the Author)
Thank you very much for the opportunity of reviewing this manuscript. In the manuscript the authors claim, and present enough evidence to demonstrate monolithic integration of junction-less SOI devices performing with current densities above 500 μA/μm and with a subthreshold swing of ~80-100 mV/dec. More importantly these devices are fabricated at BEOL compatible temperatures (<400C) and with rather good reproducibility by transfer a thinned down Si layer to a target "device" substrate.
The originality of the work relies on the transfer of the already thinned down Si layer by detaching it from the "growth" wafer. This allows for low temperature budge as opposed to growing the active layer directly on the target substrate and then thin it down. Furthermore, the process can be done through roll-to-roll transfer of films and subsequently integration.

I will recommend this paper for publication after major comments are addressed:

1 – Devices shown are above 180nm channel lengths. Is there path for scaling further? Will the ~10nm film be thin enough to still guarantee good electrostatic control? Any comments?

2 – The variability of the resulting devices was presented as one of the key differentiators with respect to other technologies (oxide semiconductors, 2D materials, etc…) will it be possible to plot a benchmark where some clear metrics are shown corroborating the claim of lower variability with this process as opposed the alternatives? What about the variability of Si impact on the energy dispersion of the semiconductor?

3 – The overdrive voltage use in the devices are considerably out of range for the functional devices (~0.6VDD). Would you care to comment?

Additionally, through the manuscript clear data and methodology were presented in line with the claims of the introduction. Proper use of statistical methods and techniques was also evident and proper use of the literature available.

Referee #3

(Remarks to the Author)
The authors present a novel method for layer transfer of crystalline Si layer from an SOI wafer to a receiving Si wafer. The method involves wafer-scale roll transfer printing, where the buried oxide of SOI is etched while the released Si layer (nanomembrane) is transferred to the receiving wafer with thermal-release tape. Utilizing this transfer process, multi-layer (up to 3) 3D monolithically integrated junctionless n and pMOS transistors were fabricated. The 3D stacked transistors are also connected to form 3D circuits, demonstrating inverter, NOR, NAND, and 6T SRAM cells. The transistors are shown to have drive currents up to 650 μA/μm (for pMOS). The entire process is fabricated with temperature limited to 400°C.
The presented work has certain novel aspects, such as nanomembrane transfer with roll-transfer and the possibility to transfer onto substrates with topography. This is one of the key highlights of the work as compared to other layer transfer techniques used in Si 3D monolithic integration or SOI substrate fabrication reports.
The demonstration of transistors with good performance (for academic lab fabrication) proves that the crystalline quality of the Si layer is preserved post-transfer, and exhaustive measurements are included to show this for multiple layers. However, the proposed method as an approach for 3D monolithic Si CMOS technology has several issues from a technological standpoint as listed below:
- Area loss due to etch holes and requirement of layer alignment due to the presence of etch holes: The demonstrated etch hole diameter is 5 μm. From an advanced CMOS technology standpoint, this is a large area loss (typical metal 1 pitch in 28 nm CMOS technology is 100 nm). Even with a 90 μm hole pitch, there is a significant area within the chip that is lost. As the layer transfer is done through a mechanical stage, the hole-to-hole alignment (between layers) will be similar to hybrid bonding (https://www.nature.com/articles/s44287-024-00019-8) and hence does not offer any advantage over it. On the other hand, direct layer transfer through wafer bonding has no area loss (as there are no etch holes) and does not require any alignment between layers (DOI: 10.1109/ICICDT56182.2022.9933105). The direct bonding of SiCN-SiCN (inter-layer dielectric used in advanced CMOS) can be achieved at 250°C.
- Lack of reusability of SOI and associated cost: As the sacrificial oxide in the proposed layer transfer is actually the buried oxide of the SOI wafer, every layer transfer requires a fresh SOI wafer. This is not a cost-effective method for advanced CMOS technology. Hence, for Si CMOS 3D monolithic integration, SmartCut process is preferred (https://ieeexplore.ieee.org/document/10413807). The authors mention that the advantage proposed in the current work is that the Si nanomembrane thickness is defined by oxidation of the top Si in SOI and not polishing. This method is followed in almost all Si CMOS 3D monolithic works with direct wafer bonding/SmartCut. The polishing is only performed to remove defects due to SmartCut process and is preferred due to recyclability. Si layer thinning with oxidation is a standard process in SOI CMOS technology.
- Lack of detailed bond strength and defectivity study: There are several studies of Si-SiO2 bond strength and defectivity levels necessary from CMOS technology standpoint. Typically, the defectivity is measured with a scanning acoustic microscope or IR imaging of the bond interface. An optical image can not completely reveal the minute defects that lead to process issues in advanced CMOS fabrication. Bond strength targets (DOI: 10.1088/2043-6262/1/4/04300) are also based on requirements of advanced CMOS process (CMP, stress liners, layout-based stress,etc.). Hence, annealing is performed at elevated temperatures to attain the necessary target. It is also known that Si-SiO2 interface yields good strength at 200°C (DOI: 10.1088/2043-6262/1/4/04300). The choices made in advanced Si 3D monolithic work are based on all these considerations. The claims made in the paper regarding low temperature process, are not done under these constraints.

- Low performance of junctionless transistors: Junctionless transistors do not meet the drive current requirement of the advanced nodes, where 3D monolithic architecture is being considered as a potential scaling option. Typically, the requirements are in the order of mA/µm. The junctionless architecture suffers from high contact resistance, extension resistance, and limited threshold voltage tunability, which are key requirements in CMOS technology. The authors present 650 µA/µm for pMOS. However, this is at much higher overdrive than CMOS standard Vov=Vdd. At the standard overdrive, the drive currents are much lower. Therefore, alternative approaches are being explored: https://ieeexplore.ieee.org/document/9265026. In addition, the authors must mention 3D monolithic CMOS reports through direct wafer bonding and benchmark junctionless transistors in those reports (DOI: 10.1109/VLSIT.2018.8510705)as these are the most relevant for the proposed approach. Figure 2 should include the junctionless devices mentioned above in the benchmark and in the extended data figure 9.
The thermal budget discussion in the paper refers to older work and is only relevant for old technology nodes. Since 3D monolithic is being considered for advanced nodes, recent work has demonstrated higher temperature tolerance (https://cea.hal.science/cea-03727300/file/VLSI2022_XGARROSv2.pdf). All of these factors need to be considered for CMOS technology.
- The authors should present drain induced barrier lowering (DIBL) and subthreshold slope at high Vd in Figure 2. This is a necessary characterization for short channel devices for digital applications.
- 3D circuits: The discussion surrounding 3D circuits has a few statements regarding the reduction in area compared to 2D counterparts (lines 22 and 26, page 8). The actual footprint reduction is less than what is claimed. This is understood in 3D circuits, as the overhead due to 3D via, metal placement does not provide 3 times reduction for every cell (see: DOI:10.1109/ISLPED.2017.8009189). The authors should mention it appropriately.
Although the layer transfer process is novel, the transistor and circuit demonstrations are complex and exhaustive (considering it was performed in an academic lab), the work misses important technological aspects of relevance for Si CMOS 3D monolithic integration and circuits. Therefore, it is not impactful to advance 3D monolithic CMOS technology. Several advanced demonstrations already exist (DOI: 10.1109/IEDM45741.2023.10413864), and there are reports of several generations of improvements in device architecture (including junctionless transistors), layer transfer, and bonding methods.
Highlighting demonstrations that bring out the novelty of the approach- ability to conformally transfer on topography and then demonstrate relevant devices that require such topography (as shown in: https://www.nature.com/articles/s41467-021-27208-5#Sec10) would be more appropriate than the current proposal of 3D monolithic CMOS.

Referee #4

(Remarks to the Author)
The authors reported wafer-scale roll transfer of ultrathin single-crystalline silicon nanomembranes to achieve monolithic 3D integration of complementary junctionless transistors under BEOL-compatible processing temperature (<400 °C). They further demonstrated high-performance p- and n-type junctionless FETs, inverters, NAND/NOR logic gates, and SRAM cells across three tiers. The manuscript is suitable in Nature, but there are several points that need to be addressed.

1. The authors demonstrate device performance uniformity; however, during the transfer process, significant residues can remain. When exposed to thermal treatments, these residues may induce additional variability in device performance. It would strengthen the manuscript to include thermal stability tests at 400 °C as a function of annealing time to quantify how transfer-induced residues affect device performance.

2. The author demonstrated multi-tier devices using silicon nanomembrane junctionless transistors. Even though the transistors bypass the need for dopant activation that often requires high temperature, source/drain formation through epitaxy/implantation/annealing is still necessary to obtain low contact resistance. (In 2015, CEA Leti had demonstrated 3D VLSI with CoolCube process, in which solid phase epitaxy and nanosecond laser annealing are performed. VLSI IEEE, 2015. p. T48-T49.) Therefore, the reviewer is wondering whether the doping process can be performed before the transfer of silicon nanomembrane to define the source/drain of the transistors.

3. TLM measurement tends to under/over-estimate the contact resistance during linear fitting of the data. The author should indicate the error bars of a set of data rather than fitting values extracted from a single device.

4. Vdd of 3V is considered too high for a silicon CMOS circuit. Suitable Vdd for a silicon device at legacy node typically ranging from 0.9V – 2V. The reviewer suggests the author to show the voltage-transfer characteristics of the devices in suitable Vdd.

5. The reported p-type junctionless Si FETs deliver higher current density than their n-type counterparts. While this might seem counterintuitive based on the typical electron vs hole mobility in silicon. The authors should provide a more detailed discussion of these mechanisms to clarify why the p-FETs outperform the n-FETs in this work.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)

I have carefully reviewed the revised version of the manuscript and the authors' response to the previous review comments. The authors have made substantial efforts to address the major concerns raised in the initial review. The manuscript has improved notably in clarity, methodology, and presentation.

A. Summary of the key results: After the major revisions, the abstract is now clearly formulated and provides a concise and understandable overview of the main findings. The key results are presented in a coherent manner, improving the accessibility and impact of the manuscript.I appreciate that the authors have followed my suggestion to reduce the number of references and make them more specific.
While the abstract has improved significantly, one more sentence about the roll-to-roll transfer-printing process and it's practical relevance would make it even more complete.

B. As stated in my initial review, I consider the work to be both original and significant, addressing an important topic in the field.

C. The methodology appears sound and appropriate for the research question. The data quality is adequate, and the results are presented clearly with well-prepared diagrams and figures.

D. Appropriate use of statistics and treatment of uncertainties: The authors show a strong awareness of the statistical relevance of their measurements. In most figures, results are presented for ensembles of devices, often using boxplots or histogram-style representations. This provides a clear picture of device-to-device variability and indicates a careful treatment of uncertainties.

E. Conclusions: robustness, validity, reliability: The authors have added a sentence addressing the relatively modest performance of the devices, which helps clarify this aspect of the results.
The authors have improved the conclusion section in the revised version. The conclusions are now clearly stated and better supported by the presented results. I am satisfied with the authors' final interpretation, and this aspect of the manuscript can be considered adequately addressed.

F. Suggested improvements: experiments, data for possible revision. The major points raised in the initial review have been sufficiently addressed. I have no further requests for additional experiments or data, and no further revision is required from my side. The authors briefly mention a possible upscaling of their approach in the response letter. I find this aspect very interesting and promising, as it could strongly increase the practical relevance of the work. Although it may not be essential for the current manuscript, it represents an exciting direction for future research.

G. References: appropriate credit to previous work? The references are appropiate and reflect the state of the art quite well. One minor revision point is to explicitly mention the applied transfer technique either in the abstract and in the list of keywords. In my understanding, this technique represents the key methodological approach of the work and should therefore be highlighted more prominently.

Referee #2

(Remarks to the Author)
Good morning,

Thank you very much for the detailed reply to the questions raised in my previous evaluation. All my concerns were addressed in a clear, honest, and professional manner. I particularly appreciate the authors' inclusion of additional variability data and their transparent discussion of the underlying sources of variability linked to integration limits, as well as the correlation with semiconductor and oxide thickness.

Similarly, the response regarding channel-length scaling was thorough. The added TCAD simulations—showing a scaling limit near ~30 nm and the need for more advanced architectures beyond that point—provide valuable context and further strengthen the credibility of the work.
Overall, I am satisfied with the clarifications and the additional data provided. I recommend the manuscript for publication as it is.

Referee #3

(Remarks to the Author)
The authors have addressed some of the technical issues and provided explanation or justification through experimental data. However, some of the following points remain and need to be addressed:
1a). Area loss due to etch holes: The authors have added explanation and supporting data to show Si layer release can be achieved with low density (500 µm pitch) holes. This is satisfactory for the current status and provides readers a context.
1b) Layer transfer alignment: This point was raised previously and is a drawback of the method. The authors have also included a statement "These dead space patterns can be aligned across tiers during transfer using a mechanical stage, with micron-scale resolution". Hence, direct wafer bonding remains free of any overlay errors and better than the proposed method for high performance logic, especially when 3D circuits are needed.
1c) Comparison with direct wafer bonding: Authors claim multi-layer stacking beyond two layers as unique advantage. Direct wafer bonding (DWB) layer transfer has fundamentally no restriction as well for multi-layer transfer. More than 2 layers have

been demonstrate in hybrid bonding (which uses same mechanism as DWB with added complexity of embedded interconnects (https://doi.org/10.1109/TCPMT.2025.3533926; DOI 10.1109/ECTC51529.2024.00106). As example of only layer transfer (without embedded interconnects) 3 tier has also been demonstrated (https://doi.org/10.1109/SOI.2007.4357865). This aspect is not novelty of the proposed method. DWB and hybrid bonding demonstrations do not use thin dielectrics between layers as most of those involve advanced CMOS where inter-tier metals are needed. It is not the limitation of the method but a technological choice. Same would apply to proposed method. Any advanced CMOS demonstration would require inter-tier metallization, technologically relevant gate height and other modules that would necessitate a thicker inter-tier dielectric. Regarding the cost argument, the claims can not be fully substantiated and also compared between an academic instrument/process and 300 mm industrial grade process. Only feature that remains unique to the process compared to DWB is the ability to transfer over topography, which in the revision hasn't be expanded upon.
The authors should remove the sentences that claim the process allows more tiers compared to low-temperature DWB.
2a Authors have demonstrated layer transfer for bulk wafer with their method. Unfortunately, it adds lithography layers which is adds cost and is leads additional overlay requirement. While the data is necessary to show utilization of the method with bulk wafers, it is not equivalent to reusability as in SOI wafer which is full layer transfer without the need for lithography on donor wafer. Moreover, Si (111) wafers are not standard. Please add AFM scan showing the thickness variation across the layer transferred from bulk wafer (show the variation closer to the edge of the pattern) .
2b. The method is ok for research approaches.
2c: Donor wafer CMP is only needed for recycling and specifically the need for high volume manufacturing to reduce costs. The proposed method from bulk wafers requires lithography and adds overlay overhead and is not applicable for full layer transfer (unlike SmartCut). Hence the claimed benefits are not justified. Oxidation based thinning of Si layer is fairly standard in SOI CMOS flow and is also used for SmartCut like process. Non-uniformities, contamination risks also exist in oxidation based process as well. Since the top-Si thickness variations and contamination levels across SmartCut SOI wafers is the same as standard bulk wafers (both being used in high-volume CMOS manufacturing), the risks claimed due to CMP are clearly not show stoppers. The authors should modify the sentences to put their work context for research and low volume prototype.
3: Defectivity characterization is sufficient
4: The authors have attempted to demonstrate junction-less flow as competitive. The issue is access and contact resistance and these can not be mitigated in FinFET geometry. In fact, they become worse. It is more challenging to obtain higher drive current in FinFET compared to planar geometry. The authors demonstrate layer transfer with doped source/drain layers. The same method can also be used in DWB method of 3D monolithic and is not unique benefit of proposed method. The approach leads to significantly overlapped transistor design (high gate-source/drain capacitance) and channel damage during doped layer etch. Hence it is not a useful process option. Post-layer transfer, doped source/drain in a self-aligned process is a necessity and is not solved by the process shown by authors. These details do not address the issue and justify the claim of high performance with low temperature.
Also, the sentences added by authors seem to indicate planar MOSFETs have limited drive current than FinFETs. This is an incorrect assumption without all parameters factored in (short-channel effects, footprint, necessary overdrive and supply voltage etc.). The authors should remove the sentences "as limited mainly by the planar channel structure. Future iterations that transition from planar channels to vertical fins are expected to significantly narrow this performance gap".
Overdrive conditions: Constant overdrive comparison is used to compare at equivalent field for various EOT. The authors have misunderstood the concern. It is expected that EOT of academic devices is lower than that of state-of-the art CMOS. However, constant overdrive comparison allows current and charge density in the channel comparison across these. The benchmark table should show ION at fixed overdrive conditions across technologies. The figure 14b is misleading since it compares ION and various conditions across technologies and should be removed.
Benchmarking: The compared Si DWB benchmark references utilizes 525°C because of the raised S/D module (RSD). As mentioned in the previous point, state-of-the art transistor architecture requires non-overlapping flow. The bonding and transfer process itself requires temperature below 400°C. This context should be mentioned.
5: DIBL is not close to simulation value. In CMOS technology terms it is a large offset 11 mV/V vs 60 mV/V. Modify the sentences. Subthreshold slope in high Vd is not mentioned.
7: Cost and manufacturability: Authors have not done any cost model comparison to claim the benefits. Particularly, the proposed layer transfer from bulk wafers adds lithography steps (that are known to be expensive and causing yield issues) and adds overlay errors. Therefore, it is not equivalent to wafer-bonding layer transfer.
Complementarity: While the proposed process allows layer-transfer on topography, and relevant for research (as mentioned by authors), it should reflect in the abstract and main claim of the paper.
The proposed work's novelty is in transfer process. The subsequent MOSFET process flow does not feature novel process or conditions that could make them high performance. Low temperature process (400°C) is due to the choice of the process flow that does not include process modules necessary for higher performance CMOS. The challenges for <=400°C are due to various CMOS process modules (spacers, RSD, high-k/metal gate etc.) which are integral part of advanced CMOS flow (authors can refer to extensive studies done on Si 3D Monolithic integration with DWB).
Authors could have improved the manuscript by adding relevant device data (based on high topography features) to demonstrate the clear benefits of the proposed layer transfer.

Referee #4

(Remarks to the Author)
Overall, the responses are clear and technically sound, and the revisions strengthen the manuscript without introducing new ambiguities. I therefore recommend acceptance of the manuscript in its current form.

Referee #1:

Summary comments: *"A. the approach of monolithic three-dimensional integration of complementary single-crystalline silicon layers is new and original."*

Our response: We sincerely thank the referee for the positive assessment and for recognizing our approach to monolithic three-dimensional integration of complementary single-crystalline silicon layers as "*new and original*." We greatly appreciate this encouraging evaluation. In the following responses, we address each of the technical concerns raised and have revised the manuscript accordingly to further strengthen the clarity, rigor, and impact of our work.

Comment #1: *"The main issue is, the authors have stated that the method of layer transfer introduced here would lead to higher integration density of devices. However, in the device section they have shown that one device layer (tier) has only 625 transistors on an area of 40 x 40 mm2, which is extremely low. One can only speculate what the reason is for that low integration density but I suspect the release holes cause the problems. You have to put the transistor in an area far away from the release hole? But even if circuit designers would be able to use a design kit for those transistors I cannot see how to deal with the holes. Is the size of a circuit limited to 85 x 85 µm2? In my opinion it is too small for VLSI."*

Our response to comment #1: We appreciate the referee's thoughtful question regarding integration density and the role of release holes. We clarify below that the demonstrated array in Fig. 3 is not a density limit, that release holes do not constrain circuit size, and that the area overhead is negligible, in four parts:

First*, *the density (625 devices over 40 × 40 mm^2) demonstrated in Fig. 3 does not represent the achievable integration density: The sparse layout shown in Fig. 3 was not intended to reflect the maximum device density of our platform. Instead, it was designed to map device spatial uniformity and yield across the entire transferred film on wafer scale. We therefore deliberately distributed transistors very sparsely for the purpose of keeping the total device count manageable for measurement using a manual probe station. Thus, the demonstrated 625-device array is not a density demonstration; it is a metrology structure. The achievable integration density is many orders of magnitude higher, as illustrated in Fig. 2c. We have clarified this point in the revised manuscript.

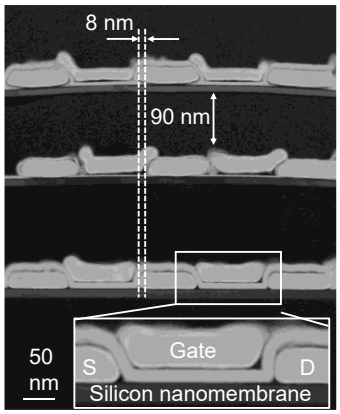
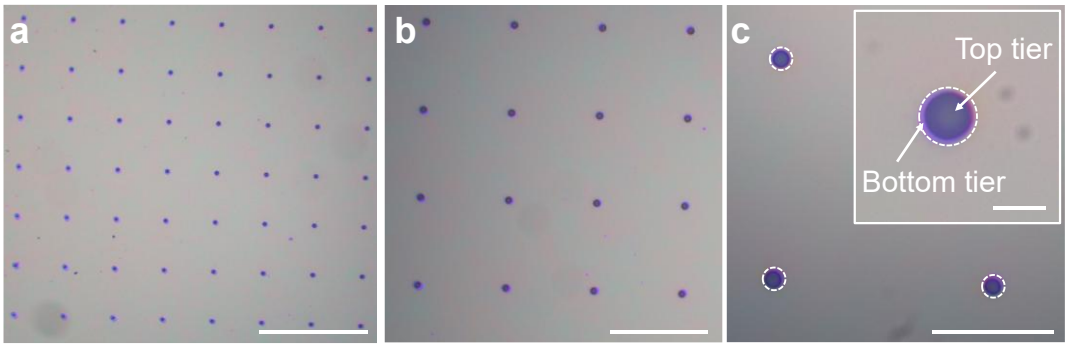


Figure 2. BEOL-compatible junctionless transistors built on transferred single-crystalline silicon nanomembranes. **c**, STEM micrograph showing three tiers of stacked junctionless transistor arrays separated by ~90 nm low-κ interlayer dielectrics. Scale bar: 50 nm. Inset: Zoom-in view of a single device.

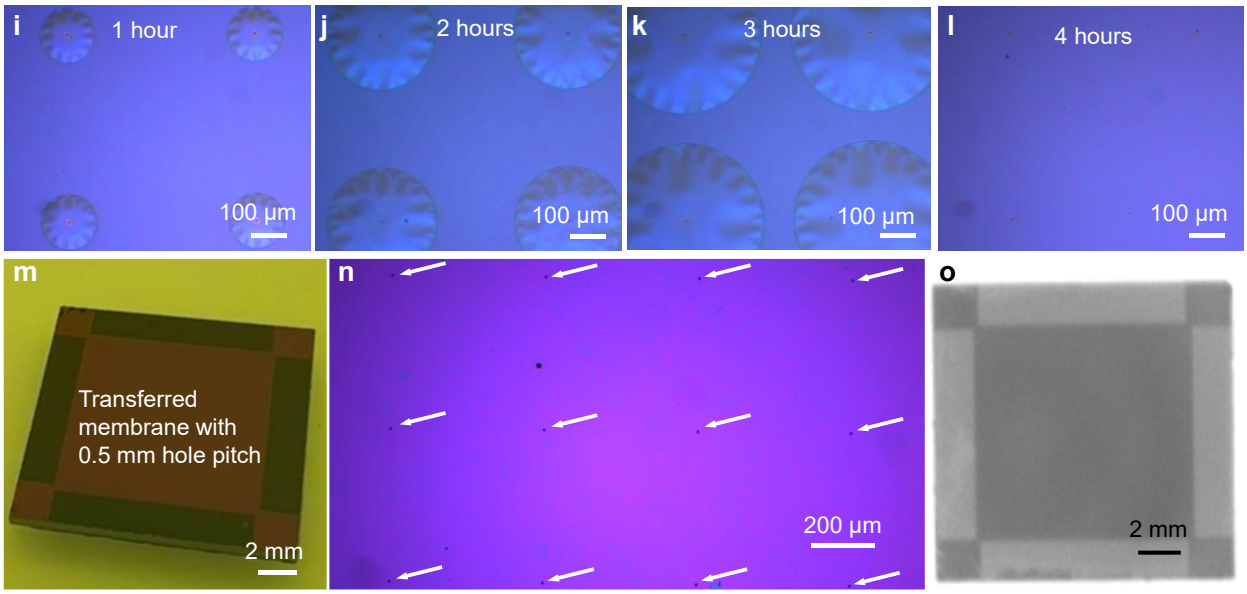
Second*, *release holes do not limit circuit size. Since the locations of release holes are known a prior, they are simply excluded through standard design-rule constraints in the process design kit (PDK), exactly as what is done in our device and circuit demonstrations, including the device arrays covering 40×40 mm^2. This is the standard practice in IC design, where keep-out regions, e.g., TSV, alignment marks, dummy fill, are automatically handled by the PDK. Thus, the presence of release holes does not constrain circuit size to 85×85 μm^2 or any similar limit and the designers do not need to deal with them manually.

Third, transistors do not need to be placed "far away" from release holes. A margin of ~1 μm is sufficient to accommodate tier-to-tier alignment, as shown in Extended Data Fig. 7.



Extended Data Figure 7| Alignment of sacrificial access patterns between different tiers of transferred single-crystalline silicon nanomembranes. Optical images showing the misalignment less than 1 μm under low (part **a**, scale bar: 100 μm), medium (part **b**, scale bar: 50 μm), and high (part **c**, scale bar: 25 μm) magnifications. Inset: zoom-in view of the access patterns, scale bar: 5 μm. The white dashed circles as visual guide highlight the boundary of keep-out regions as part of the design rule in the PDK to avoid these holes.

Fourth, area loss caused by release holes is small and scalable to a negligible level of <0.01%. To further address the referee's concern, our quantitative analysis shows that the area overhead introduced by accommodating these release holes in PDK is extremely small. Take the structure adopted in Extended Data Fig. 7 shown above as an example, with a 5 μm-diameter keep-out region around each hole on a 50 μm pitch, the fractional chip-area loss in PDK is only $\pi \cdot 2.5^2/50^2$=0.8%. This overhead is far smaller than the >2× density gain enabled by monolithic 3D stacking. To further substantiate this point, we added new experiments demonstrating successful release and transfer with hole pitches up to 0.5 mm. At this pitch, the area overhead becomes $\pi \cdot 2.5^2/500^2$=0.008%, which is effectively zero and has virtually no impact to device integration density. Crucially, increasing the pitch does not degrade undercut etching or transfer quality, as confirmed by optical and IR imaging (Extended Data Fig. 3i–o).



Extended Data Figure 3| Smooth undercut etching of buried oxide by HF with the help of triton surfactant. **i-l**, Optical micrographs showing the progress of oxide undercut etching as a function of time from 1 hour (part **i**), 2 hours (part **j**), 3 hours (part **k**), to 4 hours (part **l**) for 0.5 mm pitch. **m-n**, Optical image of the silicon nanomembrane with 0.5 mm hole pitch transferred to a SiO_2/Si substrate (part **m**), with zoom-in view (part **n**). Arrows serve as visual guide to mark the positions of release holes. **o**, Infrared image of the same sample confirms that the increase of pitch does not affect the quality of the buried interface with ILD.

Our modification to the manuscript: We added the statement in the revised manuscript to clarify the integration density: "*For this demonstration, the transistors were deliberately spaced at low density to enable mapping of film uniformity and device yield across the transferred layers, whereas the achievable integration density is substantially higher, as illustrated in Fig. 2c.*"

We also added the following statement and new data shown above in Extended Data Fig. 3 to clarify the area loss: "*The regions where silicon was removed to provide HF access to the sacrificial buried oxide account for only a minor area loss. Quantitatively, with a hole diameter of 5 μm on a 25 μm pitch (Fig. 1c and 1d), the fractional area loss is merely* $\pi \times 2.5^2/25^2 = 3.1\%$. *This overhead can be further reduced by scaling the hole diameter and pitch. As shown in Extended Data Fig. 3, increasing the pitch to 0.5 mm does not affect the transfer process, while reducing the area loss to only* $\pi \times 2.5^2/500^2 = 0.008\%$, *which is effectively negligible. In practice, such regions are excluded by design rules in the process design kit (PDK). These dead space patterns can be aligned across tiers during transfer using a mechanical stage, with micron-scale resolution (Extended Data Fig. 7).*"

Comment #2: *"Another disadvantage is the preparation of the silicon layers for the transfer. You have to sacrifice one SOI wafer for each Si sheet you want to transfer. When stacking it as suggested with many tiers the cost and the effort will be enormous."*

Our response to comment #2: We thank the referee for raising this important point. We agree that relying exclusively on SOI wafers to generate silicon nanomembranes would increase

cost if each transferred layer required sacrificing an SOI substrate. We address this concern in two parts. **Importantly, we do not need to sacrifice SOI wafers with a new reusable bulk-wafer method.**

First, in the near term, monolithic 3D integration is targeted toward applications with lower cost sensitivity. Our approach is initially aimed at high-performance computing and defense systems where SWaP (size, weight, and power) metrics are critically important. For these use cases, cost sensitivity is lower and monolithic 3D integration offers substantial performance benefits. This situation is analogous to the adoption of 3D high-bandwidth memory (HBM) in AI accelerators, where the performance gains justify the higher wafer-level cost. Nonetheless, we fully agree that long-term scalability requires a more cost-effective source of ultrathin single-crystalline silicon membranes.

Second, new experimental results demonstrate that silicon membranes can be derived from low-cost, potentially reusable bulk silicon wafers. Meanwhile, to directly address the referee's concern, we have also added new experiments demonstrating that freestanding silicon nanomembranes can be prepared from reusable bulk (111) silicon wafers, thereby eliminating the need to sacrifice SOI wafers. The process is summarized in Extended Data Fig. 4 and detailed in Method. Briefly, we started with a bulk (111) silicon wafer and formed the release structures as interlocking lines perpendicular to the <110> direction. The top surface of the wafer and the sidewalls of these interlocking release trenches are then protected by a thermal oxide and nitride mask. Highly anisotropic undercut etching along <110> direction releases a silicon nanomembrane that can be roll-transferred using the same method as in our SOI-based demonstration. After membrane release, the bulk wafer substrate remains intact and can be repolished and reused, substantially reducing cost. The transferred membranes exhibit no detectable voids or defects, as confirmed by optical, SEM, and IR imaging (Extended Data Fig. 4g–i).

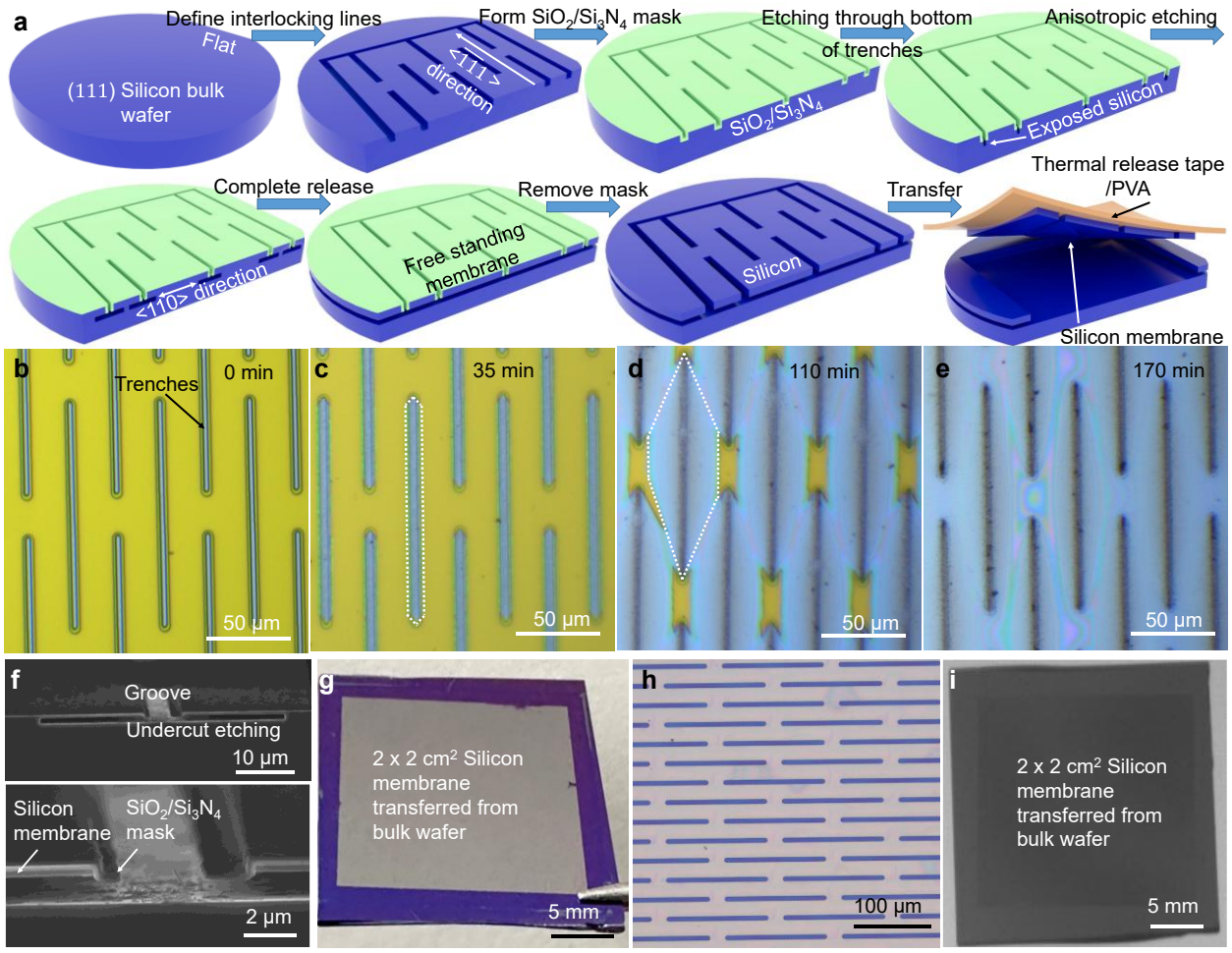
These results establish the feasibility of a cost-effective, reusable bulk-wafer process, and we have incorporated this discussion and the new data into the revised manuscript. Further optimization of this bulk-wafer-based approach will be pursued in follow-up studies.

Our modification to the manuscript: We added the following statement to the revised manuscript: "*In addition to utilizing SOI wafers, freestanding silicon nanomembranes can also be derived from bulk silicon (111) wafers through anisotropic etching, providing a potentially lower-cost fabrication route (See Method and Extended Data Fig. 4)*[37]."

Method

Fabrication of freestanding single-crystalline silicon nanomembranes from bulk silicon wafers. The process is schematically illustrated in Extended Data Fig. 4a. A (111)-oriented silicon wafer (Prime grade, University Wafers) was first spin-coated with AZ5214E photoresist and patterned into interlocking lines, which were oriented perpendicular to the $<\bar{1}10>$ crystallographic direction. Using the photoresist as an etch mask, CF_4 RIE produced arrays of grooves, with the etching depth defining the eventual nanomembrane thickness. A 200 nm thermal oxide was then grown by dry oxidation to cover the silicon surface, followed with plasma-enhanced chemical vapor deposition of 170 nm Si_3N_4, which collectively served as the hard mask for subsequent etching. A second photolithography step defined openings within the grooves. RIE was then used to sequentially remove the oxide-nitride mask at the groove bottoms

and etch the underlying silicon, producing ~700 nm deep trenches. After stripping the photoresist in hot acetone, the wafer was immersed in a preferential silicon etchant (PSE-200, Transene) at 55 ºC for anisotropic etching along the <110> and <$\bar{1}$10> directions, with the trench sidewalls protected by the SiO_2/ Si_3N_4 bilayer mask. This undercut etching released freestanding silicon nanomembranes (Extended Data Fig. 4b-e), with the high etching anisotropy confirmed with cross-sectional SEM (Extended Data Fig. 4f). Finally, etching in BOE removed the SiO_2/Si_3N_4 mask and completed the fabrication of freestanding silicon nanomembranes from bulk silicon wafers. These membranes can then be transferred using the same roll transfer-printing process to receiving substrates for monolithic 3D stacking. Optical picture, top-view micrograph, and IR image (Extended Data Fig. 4g-i) confirms the absence of defects or voids in transferred film sitting on a silicon wafer covered with spin-cast low-κ oxide ILD.



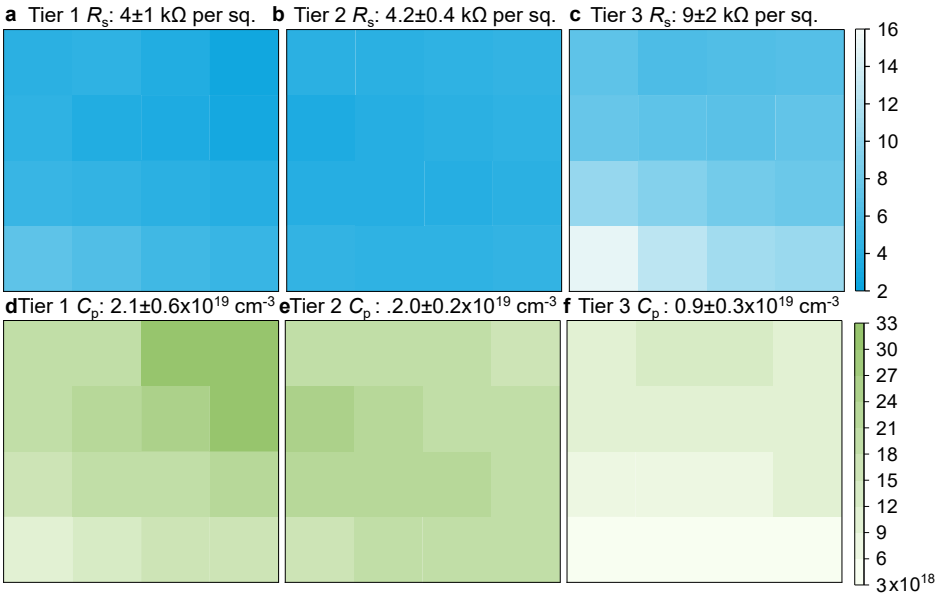
Extended Data Figure 4| Preparation of freestanding silicon nanomembranes from bulk silicon (111) wafers. a, Schematic illustration of the process flow. **b-e**, Optical micrographs showing the progress of <110> silicon undercut etching as a function of time from 0 min (part a), 35 min (part b), 110 min (part c), to 170 min (part d) at which point the silicon nanomembrane was fully released, for 50 μm pitch between 5 μm wide interlocking access trenches. White dashed lines serve as visual guide to mark the advancing undercut etching frontier. **f**, Cross-sectional SEM micrograph showing the suspended silicon nanomembrane following anisotropic undercut etching. **g-i**, Optical photo (part **g**), top-view optical micrograph (part **h**), and infrared

image (part **i**) of a nanomembrane derived from a bulk (111) silicon wafer and transferred to a SiO_2/Si subtracted covered by spin-cast low-κ ILD.

Comment #3: *"To perform the "uniform" doping by diffusion, high consumption of energy is required. Somewhere in the text, the authors claim that the doping is uniform. However, later on it is stated that the variations in threshold voltage and sub threshold swing might be caused in doping inhomogeneity in the layers. An appropriate testing the doping homogeneity would be a four point probe measurement across the wafer. A mapping would answer the question if it is homogenous or not."*

Our response to comment #3: We thank the referee for this constructive suggestion. Our use of the term "*uniform doping*" refers to the design of junctionless transistors, which employ a uniformly doped channel at the device scale (micron to nanometer) without any p–n junctions or intentional lateral doping gradients. This is distinct from wafer-scale spatial variation.

To directly address the referee's concern regarding wafer-scale doping uniformity, we included the four-point-probe mapping of sheet resistance across all three silicon nanomembrane layers prior to transfer and stacking, as recommended. We also converted sheet resistance to active phosphorus concentration using the well-established calibration from Thurber (NBS Special Publication 400-64, *The Relationship Between Resistivity and Dopant Density for Phosphorus-and Boron-Doped Silicon*, 1981). The results, presented in added Extended Data Fig. 19, provide a quantitative wafer-scale assessment of doping uniformity.

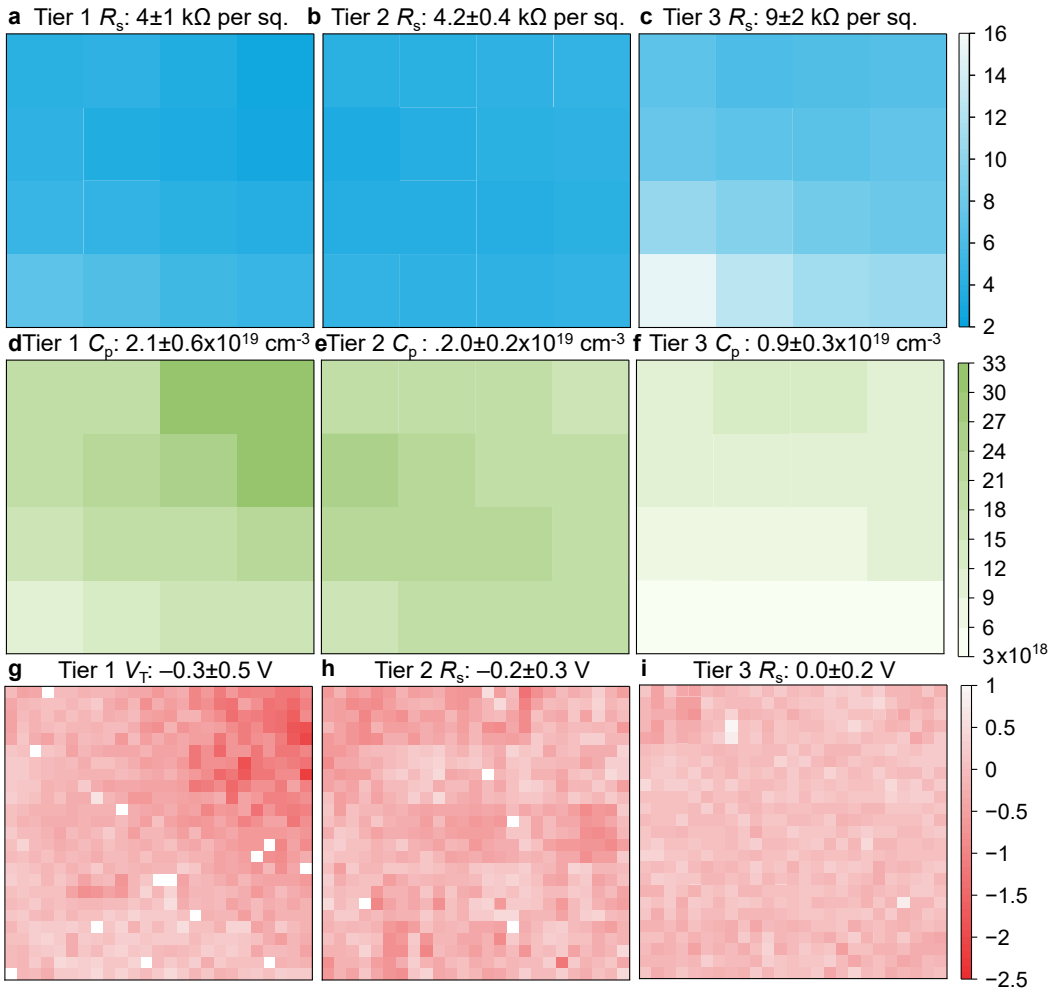


Extended Data Figure 19| Doping uniformity and its impact on device V_T. a-f, Spatial mapping of sheet resistance (R_s) measured by four-point probe and the corresponding phosphorus doping concentration ($C_{p)}$ determined from the extracted resistivity across the wafer-scale substrate for silicon nanomembranes transferred as vertically stacked tier 1 (part **a** and **d**), tier 2 (part **b** and **e**), and tier 3 (part **c** and **f**), respectively.

The mapping results confirm a reasonable degree of spatial homogeneity, with relative standard deviations in the range of 10–30% across the wafer for each doped film. We attribute the observed spatial and run-to-run variations primarily to temperature non-uniformity in the

rapid thermal annealing (RTA) furnace in our academic cleanroom. These fluctuations are not fundamental limitations of our approach and can be substantially reduced using industrial-grade diffusion or ion-implantation tools with tighter thermal and dose controls. These new measurements and the associated discussion have been incorporated into the revised manuscript.

Our modification to the manuscript: In the revised manuscript, we added: "*Further reductions of non-uniformity are achievable through improved process control; for example, tighter temperature uniformity during diffusion doping would enhance dopant homogeneity, which currently exhibits a standard deviation of 10–30% (Extended Data Fig. 19), and thereby reduce systematic V_T variation.*"



Extended Data Figure 19| Doping uniformity and its impact on device V_T. a-f, Spatial mapping of sheet resistance (R_s) measured by four-point probe and the corresponding phosphorus doping concentration ($C_{p)}$ determined from the extracted resistivity across the wafer-scale substrate for silicon nanomembranes transferred as vertically stacked tier 1 (part **a** and **d**), tier 2 (part **b** and **e**), and tier 3 (part **c** and **f**), respectively. **g-i,** Spatial mapping of V_T across the wafer-scale substrate extracted from 625 junctionless transistors on tier 1 (part **g**), tier 2 (part **h**), and tier 3 (part **i**), respectively.

Comment #4: *"The authors show statistics on their devices. The device variability across the wafers is too big. Another point not discussed, is that for lower tiers, the threshold voltage seems to change and also the variation across the wafer is increasing."*

Our response to comment #4: We thank the referee for raising this important point. We agree that the variability is larger than ideal, and we now explicitly state this limitation in the manuscript. Importantly, our measurements show that the variability is fully understood and arises from process non-uniformities in our academic fabrication environment, rather than from limitations of the silicon nanomembranes or the stacking process.

1. Origin of variability: Our metrology results indicate that the observed variability is dominated by process non-uniformities inherent to our academic fabrication environment, rather than by limitations of the silicon nanomembranes or the stacking process. In particular, as discussed in our response to the previous Comment #3, temperature gradients in the RTA furnace introduce spatially varying dopant activation, which directly leads to threshold-voltage variations with long-range spatial correlation. This effect is most pronounced in the bottom tier (Fig. S1). Specifically, four-point-probe measurements show that the phosphorus concentration increases diagonally across the wafer—from the bottom-left to the top-right—resulting in a corresponding spatial shift of V_T toward more negative values along the same direction.

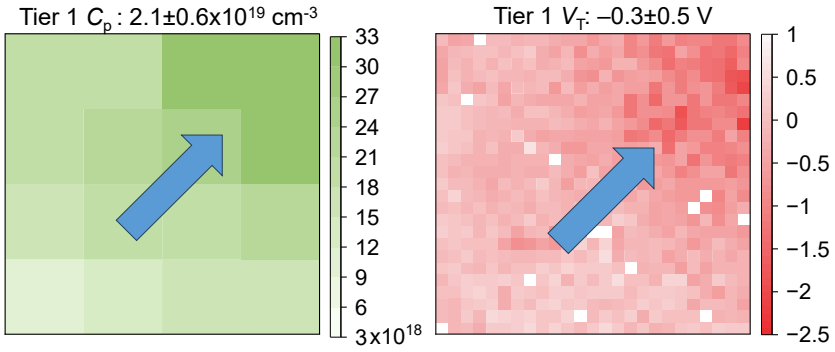
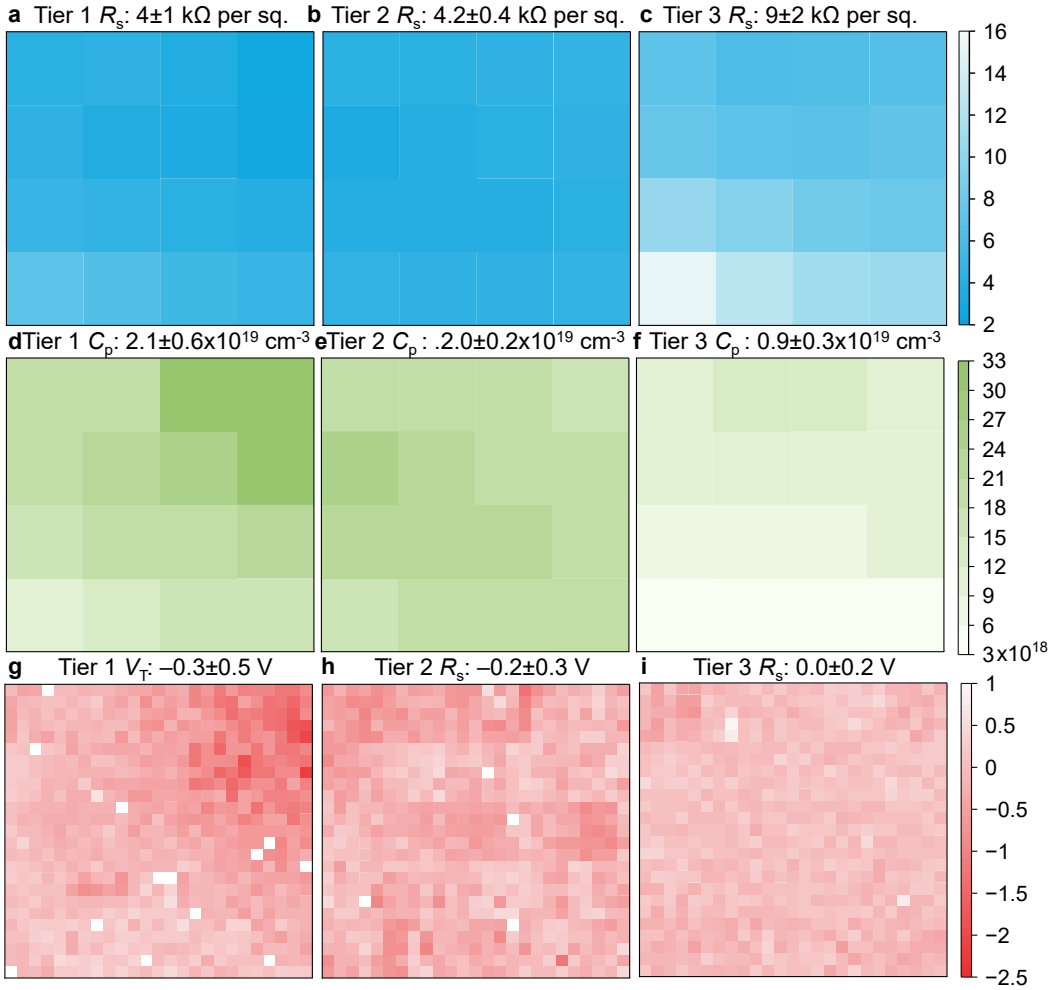


Fig. S1. Correlation between phosphorus doping concentration C_p and threshold voltage V_T. The blue arrows serve as visual guide to indicate the systematic spatial trend.

2. Explanation of difference in V_T and variability across tiers: For the top tiers, as shown in Extended Data Fig. 19, stochastic run-to-run temperature variability in the RTA results in a lower non-uniformity, with the standard deviations of phosphorus doping concentration kept at 2–3×10^{18} cm^{-3} *versus* 6×10^{18} cm^{-3} for the bottom tier. The reduced doping variation in the upper tiers before transfer directly translate to their lower threshold voltage variability (0.2–0.3 V *versus* 0.5 V in the bottom tier). Additionally, the average doping concentration is also lower in the top tier (~0.9 × 10^{19} cm^{-3}) than in the bottom tiers (2–2.1 × 10^{19} cm^{-3}), leading to a correspondingly higher average threshold voltage (0 V *versus* –0.2 to –0.3 V). These are consistent with the referee's observation, and this behavior is fully explained by doping variation existing prior to transfer and therefore does not arise from the stacking process.

We have incorporated these new experimental results and explanations into the revised manuscript.

Our modification to the manuscript: We added the following statement in the revised manuscript, in addition to the new Extended Data Fig. 19: "*The variability in extracted effective mobility and V_T arise primarily from fluctuations in channel and gate-oxide thicknesses, and most importantly from the dopant concentration[70]. Extended Data Fig. 19 shows clear systematic spatial correlations between the mean and standard deviation of phosphorus concentration and those of V_T.* "



Extended Data Figure 19| Doping uniformity and its impact on device V_T. a-f, Spatial mapping of sheet resistance (R_s) measured by four-point probe and the corresponding phosphorus doping concentration ($C_{p)}$ determined from the extracted resistivity across the wafer-scale substrate for silicon nanomembranes transferred as vertically stacked tier 1 (part **a** and **d**), tier 2 (part **b** and **e**), and tier 3 (part **c** and **f**), respectively. **g-i,** Spatial mapping of V_T across the wafer-scale substrate extracted from 625 junctionless transistors on tier 1 (part **g**), tier 2 (part **h**), and tier 3 (part **i**), respectively.

Comment #5: *"In the conclusion the authors state that their process shows good uniformity and performance like FEoL silicon MOSFETs. In FEoL MOSFETs the variation of threshold voltage is in the range of couple mV and the variation in subthreshold slope is in the range of 1-2 % across a whole 200 or 300 mm wafer. The 3D logic circuits just consists of a small number of transistors and only represent the basic circuits. To evaluate such a technique it should be appropriate to show circuits with thousands of transistor to also judge on the yield."*

Our response to comment #5: We thank the referee for this important clarification. We agree that our original conclusion overstated the comparison to FEOL silicon MOSFETs, and we have revised the manuscript accordingly.

As discussed previously, the uniformity of our BEOL-compatible transistors built on silicon, while encouraging in comparison to BEOL-compatible transistors built on 2D semiconductors and metal oxides, is limited by the various process variabilities inherent to our academic research facilities. Under these conditions, the uniformity of our current demonstration is not yet comparable to what is achievable in FEOL MOSFETs fabricated in state-of-the-art commercial silicon foundries. We have revised the manuscript to make this distinction explicit.

Although our devices exhibit a high average yield ~98–99% across tiers and from lot to lot (See our response to the next comment for more details), this level of yield is insufficient for demonstrating logic circuits comprising thousands of transistors. For example, a circuit with 1000 devices would have an effective yield of only 0.99^{1000}=0.004%. This limitation reflects the process control of our academic facility rather than any intrinsic constraint of the nanomembrane transfer or monolithic 3D stacking approach. The small-scale logic circuits included in the manuscript were intended as proof-of-concept demonstrations of functional correctness, tier-to-tier alignment, and electrical integrity, rather than as yield benchmarks.

Scaling to large-area, high-yield circuits will require industrial-grade process control. As the next step, we will work with our industry partners who have sponsored this project through NSF Industry-University Research Partnerships (IUCRC) to transfer this process to 200 mm wafers at 90 nm technology node, where the necessary uniformity and yield can be achieved.

We believe the present results merit publication on *Nature* now because they establish, for the first time, the feasibility of monolithic 3D integration of complementary single-crystal silicon transistors within a $\leqslant$400 °C thermal budget, with promising device performance, uniformity, and yield already demonstrated at the proof-of-concept scale.

Our modification to the manuscript: We added the following statement in the revised manuscript: "*Nevertheless, the device variability in our demonstration remains substantially higher than that of FEOL transistors fabricated in commercial silicon foundries, reflecting the limitations of our academic fabrication facilities.*"

We changed the statement in our conclusion into: "*Within the constraint of a BEOL-compatible processing temperature ≤ 400 °C, both p- and n-type junctionless transistors can be fabricated on transferred ultrathin silicon nanomembranes sequentially with precise sub-10 nm tier-to-tier registration, providing record-high device performance comparable to FEOL silicon MOSFETs. The observed device uniformity is primarily limited by the capabilities of our academic fabrication facilities rather than by the transfer process itself.*"

Comment #6: *"Information about the yield would be required. How is the yield on a wafer in the different tiers? How is the yield on whole wafer and also from lot to lot."*

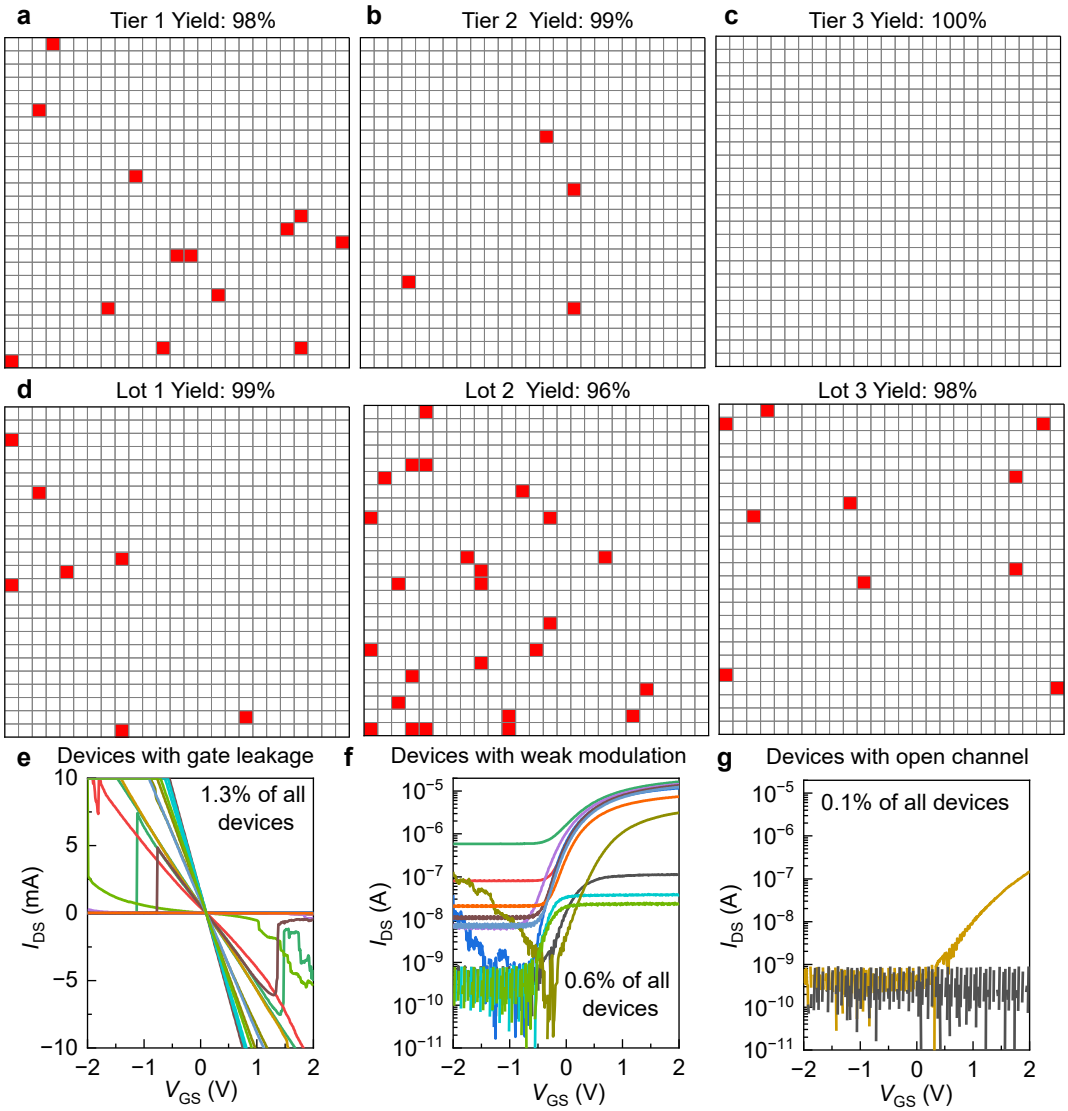
Our response to comment #6: We thank the referee for this suggestion and agree that yield information should be explicitly presented. In the experiments reported in the manuscript (Fig.

3), the device yield on a single wafer in different tiers was 98% in tier 1, 99% in tier 2, and 100% in tier 3. The whole-wafer yield, averaged across all three tiers, is therefore ~99%.

Based on reviewer's suggestion, we have fabricated three additional lots of BEOL-compatible transistors on transferred silicon nanomembranes, each containing 25 by 25=625 transistors. Across these lots, the yield ranged from 96% to 99%, with an average of ~98%. These results demonstrate that the process is reproducible across independent fabrication runs and are presented in Extended Data Fig. 22.

The small (1–4%) percentage failure cases were due to localized random defects, rather than systematic issues with the transfer or stacking process. Specifically, we observed three failure modes. 65% of failed devices (24 failures) are caused by high gate leakage current, likely due to particle contamination of surface prior to ALD growth of 10 nm HfO_2. 30% of failed devices (11 failures) are due to low or ambipolar modulation, consistent with local contamination of semiconductor-metal or dielectric interfaces. The other 5% of failed devices (2 failures) are open, likely caused by lithography defects such as broken metal lines. These failure modes are well-known in academic facilities and are not intrinsic to the monolithic 3D process. These yields support the claim that the transfer and stacking process is robust. We have incorporated these statistical results and clarifications into the revised manuscript.

Our modification to the manuscript: We added the following statement to the revised manuscript: "*The transistor yield ranges from 96% to 100%, with an average of ~98% across different tiers and from lot to lot among the 3750 devices measured. The small number of device failures are attributable to localized random defects, such as gate leakage resulting from particle contaminations, and reflect limitations of the fabrication environment of our academic facilities rather than the transfer process itself (Extended Data Fig. 22).*"



Extended Data Figure 22| Yield mapping of devices across different tiers and from lot to lot. a-c, Maps of failed devices in tier 1 (part **a**), tier 2 (part **b**), and tier 3 (part **c**) of the wafer shown in Fig. 3. **d**, Maps of failed devices from three independent lots, each fabricated on a transferred silicon nanomembrane. Failed devices are marked in red. **e-g**, Collections of transfer curves of all failed devices because of gate leakage likely caused by surface particle contamination (part **e**), weak or ambipolar modulation (part **f**), and open devices likely resulting from lithography defects (part **g**). Each colour represents a different device.

Comment #7: *"The references are adequate except the fact that the authors listed the works which dealing low dimensional nanomaterials just by naming them 10 – 22 in the abstract. But later in the table in Fig. 9 the papers are discussed in detail."*

Our response to comment #7: We thank the referee for this helpful suggestion. We agree that grouping a large number of prior works under a single block citation in the abstract reduces clarity. In the revised manuscript, we now cite carbon nanotubes and 2D transition-metal chalcogenides separately, rather than grouping them under the broad label of "low-dimensional

nanomaterials" with a block citation. The detailed discussion and quantitative benchmarking of these prior works are retained in the main text and in Extended Data Fig. 14 (Extended Data Fig. 9 in the previous manuscript), where they are presented in a more appropriate context. We believe this revision significantly improves the readability of the abstract while ensuring that the relevant literature is fully acknowledged and discussed in the body of the paper.

Our modification to the manuscript: We changed the statement in abstract into: "*BEOL-compatible transistors have been built on laser-annealed polycrystalline silicon, metal-oxide semiconductors, carbon nanotubes, and 2D transition-metal chalcogenides*[3-6]. "

Comment #8: *"There are too many papers cited in the abstract."*

Our response to comment #8: We thank the referee for this helpful comment. We agree that an excessive number of citations in the abstract detracts from its clarity. In the revised manuscript, we have reduced the references to only a few representative review papers that acknowledge prior work in the field. The complete set of relevant studies is still discussed and quantitatively compared in the main text and in Extended Data Fig. 14. We believe this revision improves the readability of the abstract while ensuring that all relevant literature is still fully covered in the body of our manuscript.

Our modification to the manuscript: We changed the statement in abstract into: "*Monolithic, three-dimensional (3D) integrated circuits promise advantages in packing density, energy consumption, and interconnectivity bandwidth but require forming high-performance semiconductors and transistors on top tiers under the constraint of a limited thermal budget compatible with back-end-of-line (BEOL) integration*[1,2]. *BEOL-compatible transistors have been built on laser-annealed polycrystalline silicon, metal-oxide semiconductors,* carbon nanotubes, and 2D transition-metal chalcogenides[3-6]."

It reduced the number of papers cited in the abstract from 22 down to 6.

Referee #2:

Summary comments: *"In the manuscript the authors claim, and present enough evidence to demonstrate monolithic integration of junction-less SOI devices performing with current densities above 500 µA/µm and with a subthreshold swing of ~80-100 mV/dec. More importantly these devices are fabricated at BEOL compatible temperatures (<400C) and with rather good reproducibility by transfer a thinned down Si layer to a target "device" substrate.*

The originality of the work relies on the transfer of the already thinned down Si layer by detaching it from the "growth" wafer. This allows for low temperature budge as opposed to growing the active layer directly on the target substrate and then thin it down. Furthermore, the process can be done through roll-to-roll transfer of films and subsequently integration.

I will recommend this paper for publication after major comments are addressed:"

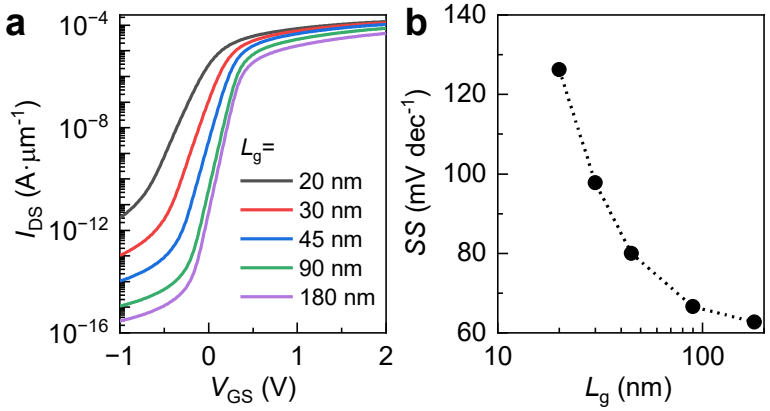
Our response: We sincerely thank the referee for the positive and encouraging assessment of our work, and for recommending the manuscript for publication in *Nature* pending satisfactory revision. We have carefully considered all comments and have substantially revised the manuscript to address each point in depth. Detailed responses are provided below.

Comment #1: *"Devices shown are above 180nm channel lengths. Is there path for scaling further? Will the ~10nm film be thin enough to still guarantee good electrostatic control? Any comments?"*

Our response to comment #1: We thank the referee for this valuable comment. We agree that the scalability of our devices should be explicitly addressed.

Junctionless transistors with ultrathin (<10 nm) fully depleted channels are known to exhibit excellent electrostatic control and strong immunity to short-channel effects, as established in prior studies (see 10.1109/LED.2011.2162309 and 10.1109/LED.2010.2099204). Our current ~8 nm thick silicon nanomembranes already ensure strong electrostatic control even in a simple planar geometry.

Beyond the experimentally demonstrated 180 nm channel length, we have added new TCAD simulations showing that our devices maintain subthreshold swing <100 mV/dec down to gate length of 30 nm, corresponding to the 32 nm technology node, even without architectural enhancements (Extended Data Fig. 10).



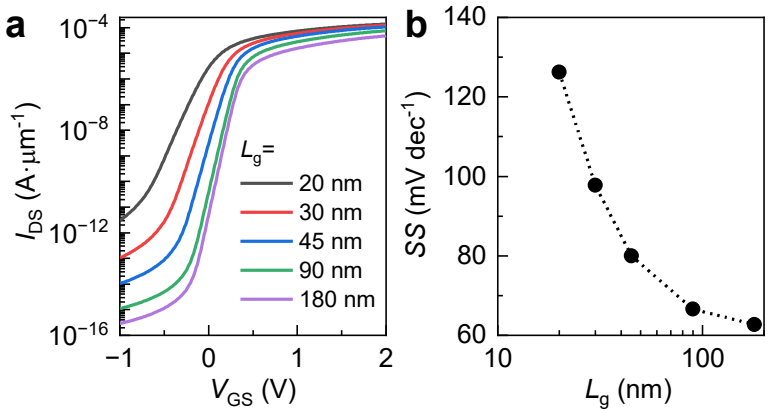
Extended Data Figure 10| TCAD simulation of BEOL-compatible junctionless transistors. **a,** Simulated transfer curves of planar junctionless transistors with a uniform channel phosphorus doping concentration of 10^{19} cm^{-3}, for gate length L_g scaled from 180 nm to 20 nm. The silicon channel thickness is 8 nm, and the gate oxide is 10 nm HfO_2. Applied V_{DS} is 50 mV. **b**, *SS* as a function of L_g.

Scaling junctionless transistors beyond the 32 nm node follows the same trajectory as advanced MOSFETs. This likely requires either adoption of fin or gate-all-around device architectures, as discussed in 10.1063/1.3079411 and 10.1109/TNANO.2025.3539457, or further reducing the silicon film thickness to ~5 nm, which has enabled sub-10 nm gate length as demonstrated in ET-SOI devices (10.1109/IEDM.2003.1269360).

The channel length of our present BEOL-compatible junctionless transistors is limited primarily by **(1)** the lithography capability available in an academic facility, and **(2)** the relatively large gate equivalent oxide thickness (EOT) arising from the SiO_2 interfacial layer beneath HfO_2. Both are process and tool related constraints rather than fundamental limitations of our approach. As the next step, in collaboration with industry partners who have sponsored this project through the NSF Industry-University Research Partnerships (IUCRC) Center for Advanced Semiconductor Chips with Accelerated Performance (ASAP), we plan to implement advanced patterning and optimized gate-stack processes at the 90 nm node and beyond.

We have incorporated these new TCAD results, and the associated discussion of scaling limits and pathway into the revised manuscript.

Our modification to the manuscript: We added the following statement regarding the scalability of junctionless transistors to the revised manuscript: "*Owing to their ultrathin (<10 nm) fully depleted channels, junctionless transistors are inherently robust against short-channel effects*[52,53]. *TCAD simulations further show that even with a simple planar geometry, they scale well to gate lengths down to 30 nm (Extended Data Fig. 10), with further improvements expected from scaling down the film thickness to 5 nm and adopting fin-type or gate-all-around architectures*[54-56]. "



Extended Data Figure 10| TCAD simulation of BEOL-compatible junctionless transistors. **a,** Simulated transfer curves of planar junctionless transistors with a uniform channel phosphorus doping concentration of 10^{19} cm^{-3}, for gate length L_g scaled from 180 nm to 20 nm. The silicon channel thickness is 8 nm, and the gate oxide is 10 nm HfO_2. Applied V_{DS} is 50 mV. **b**, *SS* as a function of L_g.

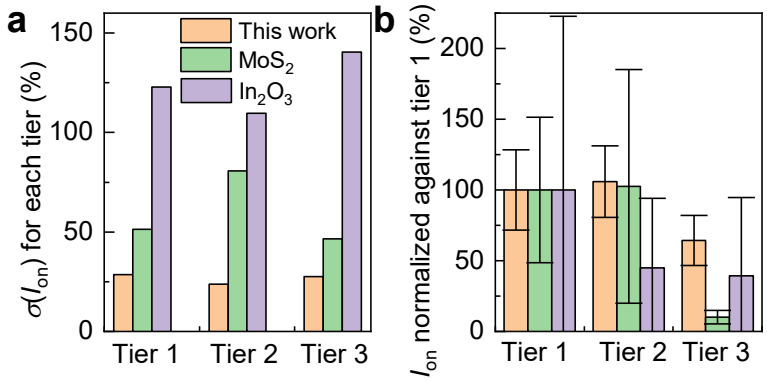
We also added the following statement in the revised manuscript to point out the importance of scaling down the gate oxide: "*Reducing the interfacial native oxide through in situ removal prior to gate oxide growth, scavenging reactions, or nitridation treatments should lower the EOT*[66,67]*, hereby reducing the required drive voltage V_{DD} and further improving device performance and scalability.* "

Comment #2: *"The variability of the resulting devices was presented as one of the key differentiators with respect to other technologies (oxide semiconductors, 2D materials, etc…) will it be possible to plot a benchmark where some clear metrics are shown corroborating the claim of lower variability with this process as opposed the alternatives? What about the variability of Si impact on the energy dispersion of the semiconductor?"*

Our response to comment #2: We thank the referee for this insightful comment. We have added a quantitative variability benchmark comparing our devices with state-of-the-art 3D integrated MoS_2 transistors reported in *Nature* (10.1038/s41586-023-06860-5) and the 3D integrated metal-oxide transistors reported in *Nature Electronics* (doi.org/10.1038/s41928-025-01469-0). We compared the variability of I_{on}, which is a clear metric capturing both threshold voltage V_T and mobility μ variations. These results confirm that our process achieves substantially lower variability than existing BEOL-compatible 3D semiconductor technologies, both within individual tiers and across stacked tiers. The quantitative results are illustrated in the new Extended Data Fig. 20.

- **Intra-tier variability**: averaged relative standard deviation of the on-state current I_{on} is 27% for our silicon junctionless transistors *versus* 60% for 3D MoS_2 transistors and 124% for 3D In_2O_3 transistors.
- **Inter-tier variability**: relative standard deviation of I_{on} is 22% for Si junctionless transistors *versus* 53% for 3D MoS_2 transistors and 34% for 3D In_2O_3 transistors.

These results confirm that our process achieves substantially lower variability than existing BEOL-compatible semiconductor technologies.



Extended Data Figure 20| Uniformity comparison among BEOL-compatible transistors. a, Standard deviation of I_{on} normalized to its average for each tier of BEOL-compatible silicon junctionless transistors (orange), MoS_2 transistors (green), and In_2O_3 (purple) transistors. **b**, Average I_{on} for each tier, normalized to the 1st tier. Error bars represent s.d..

Furthermore, we find that the variability observed in our BEOL-compatible silicon transistors is primarily due to systematic process non-uniformities inherent to our academic fabrication environment, rather than any intrinsic limitation of the ultrathin silicon nanomembrane or our stacking process based on roll transfer. Most notably, we observed a systematic shift in V_T that correlates with doping-level spatial variations (determined via four-point probe resistivity measurements), which we attribute to temperature non-uniformity of the rapid thermal annealing furnace we used for diffusion doping, as illustrated in Fig. S1. Such systematic sources of variation are expected to be significantly reduced under industrial level process controls, suggesting that the intrinsic variability of our devices could be even lower approaching that of FEOL silicon transistors.

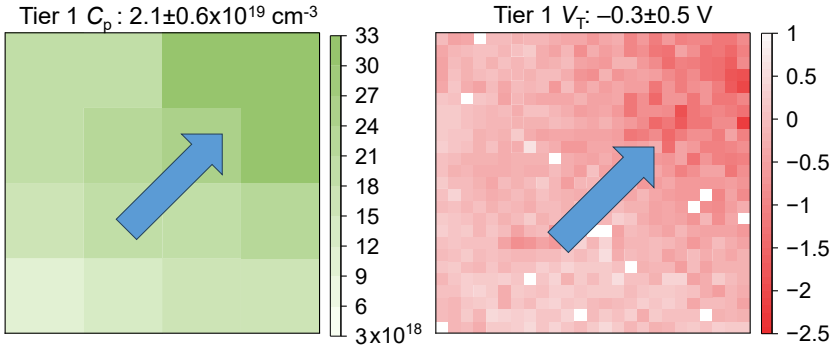
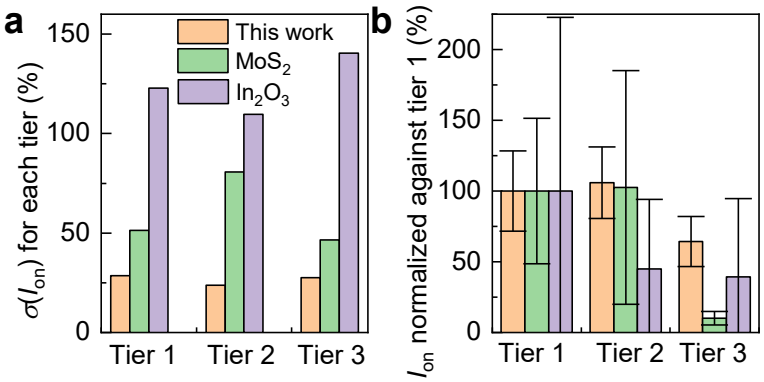


Fig. S1. Correlation between phosphorus doping concentration C_p and threshold voltage V_T. The blue arrows serve as visual guide to indicate the systematic spatial trend.

Importantly, reduced device variability directly translates into design-level benefits: a smaller guard-band envelope can be adopted in circuit and architectural design, enabling higher energy efficiency and fewer corner-case constraints.

We have incorporated the new quantitative benchmark results and the associated discussion into the revised manuscript.

Our modification to the manuscript: We added the following statement to the revised manuscript: "*The variability in extracted effective mobility and V_T arise primarily from fluctuations in channel and gate-oxide thicknesses, and most importantly from the dopant concentration*[70]*. Extended Data Fig. 19 shows clear systematic spatial correlations between the mean and standard deviation of phosphorus concentration and those of V_T. Importantly, these variabilities are substantially smaller than those reported both within individual tiers and across stacked tiers of BEOL-compatible MoS_2 and metal-oxide transistors (Extended Data Fig. 20)*[17,28]*. Nevertheless, the device variability in our demonstration remains substantially higher than that of FEOL transistors fabricated in commercial silicon foundries, reflecting the limitations of our academic fabrication facilities. Further reductions of non-uniformity are achievable through improved process control; for example, tighter temperature uniformity during diffusion doping would enhance dopant homogeneity, which currently exhibits a standard deviation of 10–30% (Extended Data Fig. 19), and thereby reduce systematic V_T variation. Lower device variability, in turn, enables the adoption of smaller guard-band envelopes in circuit and architectural design, translating directly into higher energy efficiency and fewer corner-case constraints.*"

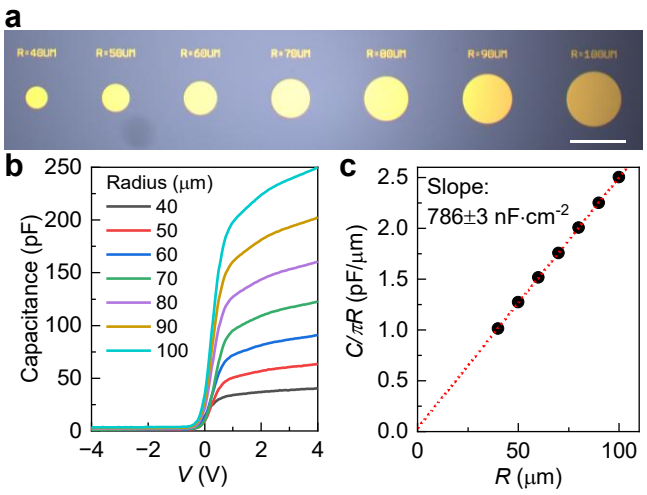


Extended Data Figure 20| Uniformity comparison among BEOL-compatible transistors. a, Standard deviation of I_{on} normalized to its average for each tier of BEOL-compatible silicon junctionless transistors (orange), MoS_2 transistors (green), and In_2O_3 (purple) transistors. **b**, Average I_{on} for each tier, normalized to the 1st tier. Error bars represent s.d..

Comment #3: *"The overdrive voltage use in the devices are considerably out of range for the functional devices (~0.6VDD). Would you care to comment?"*

Our response to comment #3: We thank the referee for raising this important point. We agree that the overdrive voltage used in our present devices lies outside the typical CMOS design envelope and we have now clarified the origin of this behavior in the revised manuscript.

The primary reason for the higher overdrive is the large equivalent oxide thickness (EOT) of our current gate stack. TSMC's 3 nm FinFET technology operates at a supply voltage of ~0.8 V as reported at ISSCC 2025 (10.1109/ISSCC49661.2025.10904560), enabled by an equivalent oxide thickness (EOT) of ~0.7 nm according to the 2023 IRDS roadmap. In contrast, our devices exhibit a more than six times larger EOT of 4.4 nm, as extracted from capacitance-voltage measurements (Extended Data Fig. 15). Since gate capacitance scales inversely with EOT, achieving comparable charge density requires a proportionally larger gate overdrive. This EOT scaling consideration accounts for why our devices are presently operated at higher supply voltages (2–4 V), corresponding to ~3–5× the values used in advanced FEOL technologies.



Extended Data Figure 15| Gate capacitance of fabricated junctionless transistors. a, Optical image of an array of circular *n*-doped silicon nanomembrane/HfO_2/Cr/Au capacitors with their radii (R) varied from 40 to 200 μm. Scale bar: 100 μm. **b**, Measured capacitance-voltage characteristics of the capacitors shown in part **a**. **c**, Capacitance (C) to πR ratio as a function of R. The red dashed line represents a linear fit to the data (black dots) to extract the capacitance per area of the native oxide plus 10 nm HfO_2 stack from its slope. The y-axis intercept gives the capacitance caused by the fringing field from the edge of electrodes.

The dominant contributor to our EOT is the ~2 nm interfacial SiO_2 layer observed in cross-sectional STEM (Fig. 1f inset). Several well-established industry techniques, such as de-oxygenated HF cleans with immediate high-k deposition (GlobalFoundries), interfacial layer scavenging (IBM, 10.3390/ma5030478), and plasma nitridation (Applied Materials, 10.1149/1.2911492), are known to substantially reduce interfacial-oxide thickness. These processes, however, require specialized equipment not currently available in our academic facility. Incorporating such techniques in future iterations should significantly reduce EOT and thereby lower the required drive voltage, bringing BEOL-compatible junctionless transistors closer to state-of-the-art FEOL operation. We emphasize that this is a process-related limitation, not an intrinsic constraint of junctionless transistors or our monolithic 3D integration method.

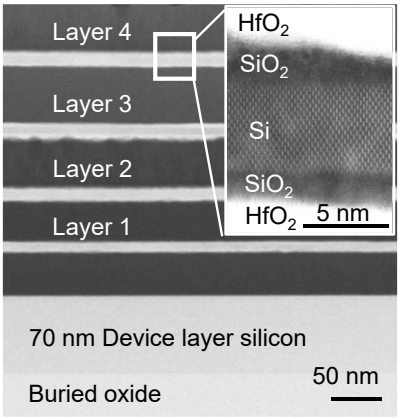


Figure 1. Wafer-scale, monolithic 3D integration of single-crystalline silicon nanomembranes. f, STEM image showing the cross-section of a stacked 5 tiers of single-crystalline silicon on a monolithic substrate. Scale bar: 50 nm. Inset: Zoom-in view of a single tier, scale bar: 5 nm.

Finally, Finally, despite the larger EOT, our 3D CMOS circuits, including inverters, NOR/NAND gates, and SRAM cells, operate successfully at supply voltages down to 0.7 V (Fig. 4). This is enabled by the sharp subthreshold swing and near-zero threshold voltage of our devices, which allow logic functionality at low VDD even before EOT optimization.

These clarifications have been added to the revised manuscript.

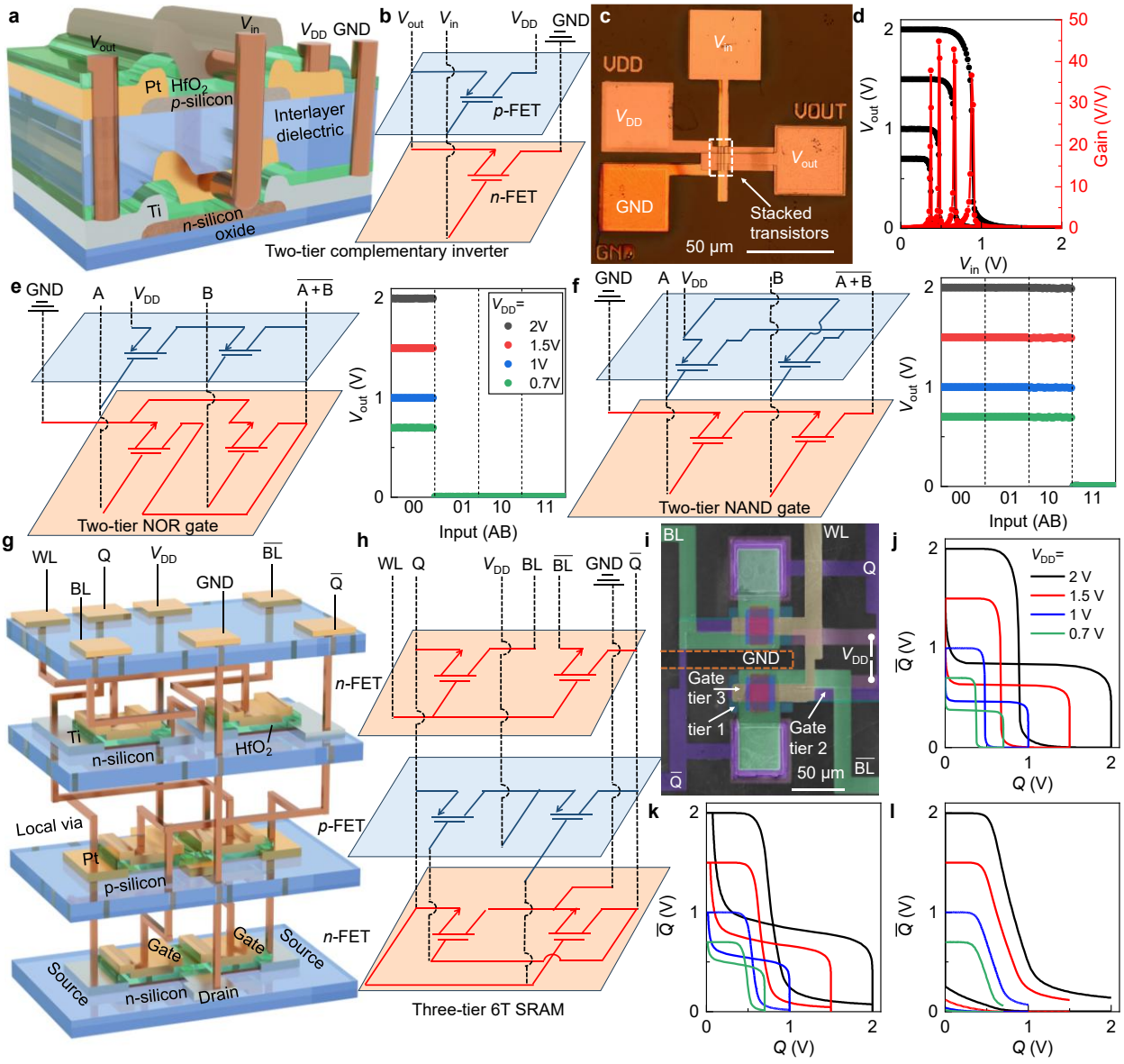


Figure 4. Monolithic 3D integrated logic circuits built on BEOL-compatible silicon junctionless transistors. a-d, Layout schematic (part **a**), circuit diagram (part **b**), optical micrograph (part **c**), and voltage-transfer characteristics (part **d**, left axis) and the associated voltage gain (right axis, red) of the monolithic 3D integrated inverter extracted with different V_{DD}. **e-f**, Circuit diagrams and output-input characteristics of a 3D NOR gate (part **e**) and a 3D NAND gate (part **f**). Applied V_{DD} ranges from 2 V (black) to 1.5 V (red), 1 V (blue), and 0.7 V (green). **g-i**, Layout schematic (part **g**), circuit diagram (part **h**), and false-coloured SEM micrograph (part **i**) of a monolithic 3D SRAM cell. **j-l**, Hold (part **j**), read (part **k**), and write (part **l**) margin measurements with V_{DD} of 2 V (black) to 1.5 V (red), 1 V (blue), and 0.7 V (green), respectively. Scale bars: 50 μm.

Our modification to the manuscript: We added the following statement in the revised manuscript: "*Our BEOL-compatible junctionless transistors employ a gate dielectric stack comprising ~1.5–2 nm native interfacial SiO_2 and 10 nm HfO_2, corresponding to an EOT of 4.4*

nm (Extended Data Fig. 15). This EOT is roughly twice that of 180 nm-node MOSFETs (EOT=2.2–3 nm)[64,65], ~2.7 times larger than 40 nm-node bonded silicon transistors[62], and more than six times greater than the ~0.7 nm EOT of state-of-the-art 3 nm technology node. Such a large EOT necessitates a higher gate overdrive voltage. Reducing the interfacial native oxide through in situ removal prior to gate oxide growth, scavenging reactions, or nitridation treatments should lower the EOT[66,67], hereby reducing the required drive voltage V_{DD} and further improving device performance and scalability."

Comment #4: *"Additionally, through the manuscript clear data and methodology were presented in line with the claims of the introduction. Proper use of statistical methods and techniques was also evident and proper use of the literature available."*

Our response to comment #4: We thank the referee for this positive remark and appreciate the recognition of the clarity of our data presentation, statistical treatment, and use of relevant literature. We are grateful for the referee's constructive feedback, which has helped us further strengthen the manuscript.

Referee #3:

Summary comments: *"The authors present a novel method for layer transfer of crystalline Si layer from an SOI wafer to a receiving Si wafer. The method involves wafer-scale roll transfer printing, where the buried oxide of SOI is etched while the released Si layer (nanomembrane) is transferred to the receiving wafer with thermal-release tape. Utilizing this transfer process, multi-layer (up to 3) 3D monolithically integrated junctionless n and pMOS transistors were fabricated. The 3D stacked transistors are also connected to form 3D circuits, demonstrating inverter, NOR, NAND, and 6T SRAM cells. The transistors are shown to have drive currents up to 650 μA/μm (for pMOS). The entire process is fabricated with temperature limited to 400°C.*

The presented work has certain novel aspects, such as nanomembrane transfer with roll-transfer and the possibility to transfer onto substrates with topography. This is one of the key highlights of the work as compared to other layer transfer techniques used in Si 3D monolithic integration or SOI substrate fabrication reports.

The demonstration of transistors with good performance (for academic lab fabrication) proves that the crystalline quality of the Si layer is preserved post-transfer, and exhaustive measurements are included to show this for multiple layers. However, the proposed method as an approach for 3D monolithic Si CMOS technology has several issues from a technological standpoint as listed below:"

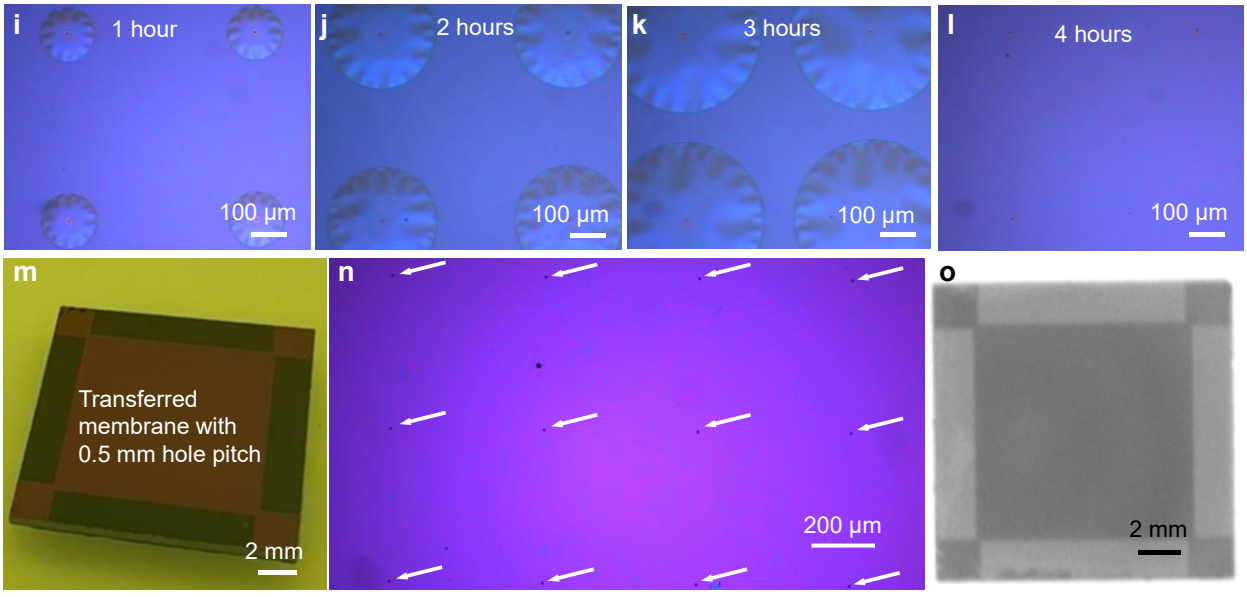
Our response: We thank the referee for recognizing the novelty of our work, including the wafer-scale roll-transfer method, tolerance to substrate topography, and the preservation of crystalline silicon quality across multiple transferred layers. We are grateful that the referee highlighted these aspects as distinguishing advantages over existing layer-transfer approaches for monolithic 3D silicon integration. We have carefully considered the technological concerns raised regarding the applicability of our method to 3D monolithic silicon CMOS integration. In the revised manuscript, we have expanded our discussion, added new experimental data, and clarified the design tradeoffs associated with our approach. Detailed, point-by-point responses to each of the referee's comments are provided below.

Comment #1: *"**Area loss due to etch holes and requirement of layer alignment due to the presence of etch holes:** The demonstrated etch hole diameter is 5 μm. From an advanced CMOS technology standpoint, this is a large area loss (typical metal 1 pitch in 28 nm CMOS technology is 100 nm). Even with a 90 μm hole pitch, there is a significant area within the chip that is lost. As the layer transfer is done through a mechanical stage, the hole-to-hole alignment (between layers) will be similar to hybrid bonding (https://www.nature.com/articles/s44287-024-00019-8) and hence does not offer any advantage over it. On the other hand, direct layer transfer through wafer bonding has no area loss (as there are no etch holes) and does not require any alignment between layers (DOI: 10.1109/ICICDT56182.2022.9933105). The direct bonding of SiCN-SiCN (inter-layer dielectric used in advanced CMOS) can be achieved at 250°C."*

Our response to comment #1: We thank the referee for raising this important point. We agree that the area overhead associated with the release-hole pattern should be quantified and compared with direct layer transfer through wafer bonding. We address the area-loss concern in three parts:

- ***First, hole dimensions are fully scalable and not a fundamental limitation.*** The release holes serve solely to provide HF access to the buried oxide, and their dimensions are scalable. The 1–5 µm diameters used in this work reflect the resolution limit of our h-line (405 nm) lithography tools rather than a process or technology constraint. With advanced lithography tools, hole diameters can be readily scaled to the nanometer regime to reduce the area loss.

- ***Second and more importantly, area loss is dominated by hole pitch, not just hole size, and is quantitatively negligible.*** The dominant factor determining the area loss is not only the hole size but the pitch between holes. In the original manuscript we demonstrated pitches up to 90 µm. In the revised version, we now include additional experiments showing scalability to at least 0.5 mm pitch (Extended Data Fig. 3i–o confirms that increasing the pitch to 0.5 mm does not affect the completeness of the buried-oxide undercut or the quality of the transferred interface). A quantitative analysis shows that the area overhead is not significant as speculated by the referee but practically negligible: with 5 µm-diameter holes on a 90 µm pitch, the relative area loss is only $\pi \cdot 2.5^2/90^2$=0.24%. At 0.5 mm pitch, the area loss is reduced to merely $\pi \cdot 2.5^2/500^2$=0.008% (See Fig. S2 for the schematic illustration). Since hole diameters can be scaled to the sub-micron or even nanometer regime with deep-UV or EUV lithography, with micron-scale alignment tolerances achievable using a mechanical stage, the 5 µm circular footprint used here represents a conservative upper bound of the area affected by each hole. Because the hole locations are known a priori, these regions can be automatically excluded as part of the design rules by the process design kit (PDK) during layout generation, exactly like TSV keep-out zones or dummy-fill regions in CMOS design.



Extended Data Figure 3| Smooth undercut etching of buried oxide by HF with the help of triton surfactant. **i-l**, Optical micrographs showing the progress of oxide undercut etching as a function of time from 1 hour (part **i**), 2 hours (part **j**), 3 hours (part **k**), to 4 hours (part **l**) for 0.5 mm pitch. **m-n**, Optical image of the silicon nanomembrane with 0.5 mm hole pitch transferred to a SiO_2/Si substrate (part **m**), with zoom-in view (part **n**). Arrows serve as visual guide to mark the positions of release holes. **o**, Infrared image of the same sample confirms that the increase of pitch does not affect the quality of the buried interface with ILD.

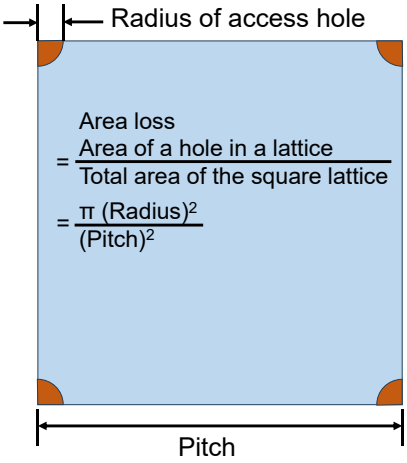


Fig. S2. Schematic of a square lattice enclosed by access holes in released silicon nanomembranes. The area loss can be calculated based on the geometric relationship.

- ***Third*, *comparison with direct transfer via wafer bonding: negligible area overhead with unique advantages.*** We acknowledge that wafer bonding can, in principle, eliminate this residual (<0.01%) area overhead. However, as the referee also noted, a distinctive advantage of our process is its high tolerance to surface topography, which enables reliable stacking of multiple layers. As shown in Fig. 1f, our method achieves vertical stacking of five crystalline silicon layers (including the base SOI substrate) within a total vertical depth of ~350 nm. To our knowledge, direct layer transfer through low-temperature wafer bonding has been demonstrated only for two stacked silicon layers (including the substrate). Moreover, our approach avoids the need for specialized, capital-intensive bonding equipment and provides a complementary integration pathway where a negligible area overhead is acceptable in exchange for:
    - Readily multilayer stacking beyond two tiers,
    - Compatibility with substrates exhibiting surface roughness or topography, and
    - Eliminating requirements of expensive wafer-bonding equipment.

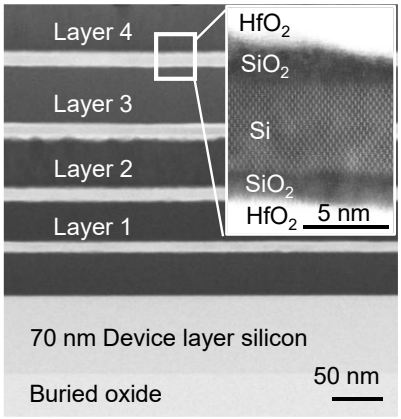
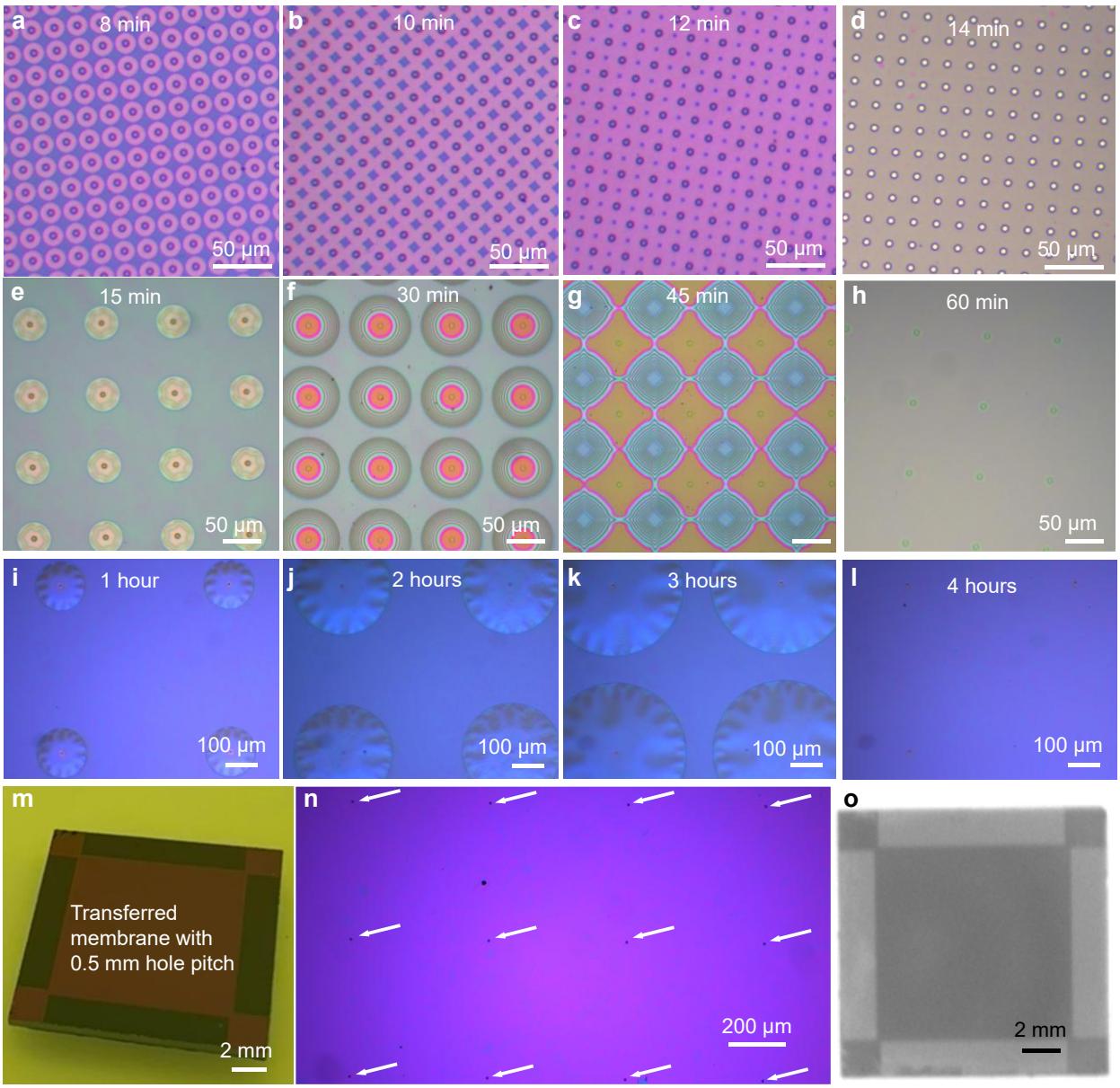


Figure 1. Wafer-scale, monolithic 3D integration of single-crystalline silicon nanomembranes. f, STEM image showing the cross-section of a stacked 5 tiers of single-crystalline silicon on a monolithic substrate. Scale bar: 50 nm. Inset: Zoom-in view of a single tier, scale bar: 5 nm.

In summary, while our approach incurs a small area overhead due to release holes, quantitative analysis shows that this overhead is negligible (<0.01%). This minor area cost is outweighed by the demonstrated ability to achieve high-yield multilayer stacking with topography tolerance and low-cost processing equipment, capabilities not currently achievable with direct-transfer via wafer bonding. We have incorporated the quantitative analysis and new 0.5 mm-pitch transfer results into the revised manuscript.

Our modification to the manuscript: We added the statement and new data to explain the area loss: "*The regions where silicon was removed to provide HF access to the sacrificial buried oxide account for only a minor area loss. Quantitatively, with a hole diameter of 5 μm on a 25 μm pitch (Fig. 1c and 1d), the fractional area loss is merely* $\pi \times 2.5^2/25^2 = 3.1\%$. *This overhead can be further reduced by scaling the hole diameter and pitch. As shown in Extended Data Fig. 3, increasing the pitch to 0.5 mm does not affect the transfer process, while reducing the area loss to only* $\pi \times 2.5^2/500^2 = 0.008\%$, *which is effectively negligible. In practice, such regions are excluded by design rules in the process design kit (PDK). These dead space patterns can be aligned across tiers during transfer using a mechanical stage, with micron-scale resolution (Extended Data Fig. 7).*"



Extended Data Figure 3| Smooth undercut etching of buried oxide by HF with the help of triton surfactant. a-d, Optical micrographs showing the progress of oxide undercut etching as a function of time from 8 min (part **a**), 10 min (part **b**), 12 min (part **c**), to 14 min (part **d**) when

the silicon nanomembrane was fully released, for 25 μm pitch between 5 μm diameter circular access patterns. **e-f**, Optical micrographs showing the progress of oxide undercut etching as a function of time from 15 min (part **e**), 30 min (part **f**), 45 min (part **g**), to 60 min (part **h**) for 90 μm pitch. **i-l**, Optical micrographs showing the progress of oxide undercut etching as a function of time from 1 hour (part **i**), 2 hours (part **j**), 3 hours (part **k**), to 4 hours (part **l**) for 0.5 mm pitch. **m-n**, Optical image of the silicon nanomembrane with 0.5 mm hole pitch transferred to a SiO_2/Si substrate (part **m**), with zoom-in view (part **n**). Arrows serve as visual guide to mark the positions of release holes. **o**, Infrared image of the same sample confirms that the increase of pitch does not affect the quality of the buried interface with ILD.

We incorporated statement discussing the benefit and limitation of our transfer process in comparison to low-temperature wafer bonding and cited the reference suggested by the referee: "*Owing to its process simplicity, transfer can be readily repeated to realize a multiple (i.e., >2) tier stack, whereas low-temperature wafer-bonding is currently limited to forming two stacked silicon layers[44], which further justifies the minimal area loss overhead introduced by the release-hole pattern. Cross-sectional scanning transmission electron microscopy (STEM) image shows an unprecedented demonstration of four additional layers of single-crystalline silicon nanomembranes added on a SOI wafer, with 40–50 nm interlayer dielectric in between (Fig. 1f).*"

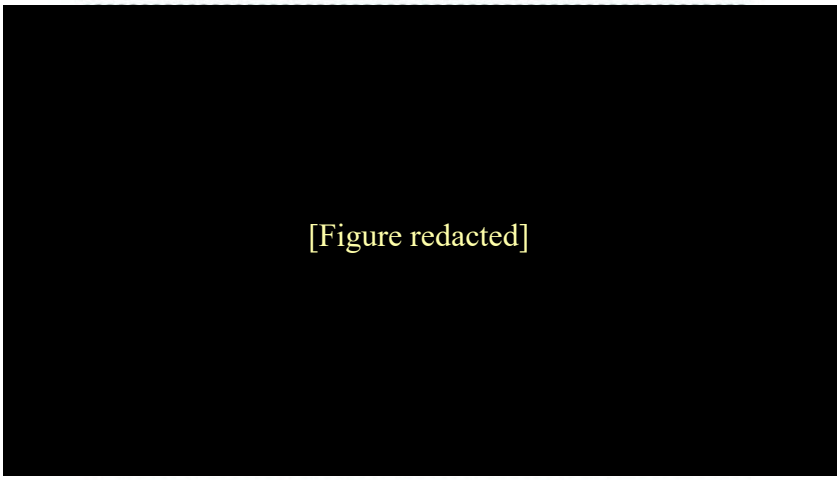
Comment #2: *"**Lack of reusability of SOI and associated cost:** As the sacrificial oxide in the proposed layer transfer is actually the buried oxide of the SOI wafer, every layer transfer requires a fresh SOI wafer. This is not a cost-effective method for advanced CMOS technology. Hence, for Si CMOS 3D monolithic integration, SmartCut process is preferred (https://ieeexplore.ieee.org/document/10413807). The authors mention that the advantage proposed in the current work is that the Si nanomembrane thickness is defined by oxidation of the top Si in SOI and not polishing. This method is followed in almost all Si CMOS 3D monolithic works with direct wafer bonding/SmartCut. The polishing is only performed to remove defects due to SmartCut process and is preferred due to recyclability. Si layer thinning with oxidation is a standard process in SOI CMOS technology."*

Our response to comment #2: We thank the referee for this important comment. We have added new results showing our process is compatible with reusable bulk wafers and we more explicitly compared our approach with direct wafer bonding/SmartCut process. Our response is organized into three parts:

First, our process is compatible with reusable bulk silicon wafers. We agree that reliance on SOI wafers in our initial demonstration could raise cost concerns. To directly address this, we have added new experimental results showing that freestanding silicon nanomembranes can also be prepared from bulk wafers, thereby mitigating reliance on SOI. Starting from a bulk (111) silicon wafer, we patterned interlocking release trenches, protected the top surface and trench sidewalls with a thermal oxide/nitride bilayer, and performed highly anisotropic undercut etching along the <110> direction. This process releases a high-quality single-crystalline silicon nanomembrane that can be roll-transferred using the same method demonstrated for SOI. After a simple polishing step, the bulk wafer can be reused. These results, now included in Extended Data Fig. 4 with the process detailed in Methods, demonstrate that our process is not intrinsically tied to SOI and can leverage lower-cost, reusable bulk silicon substrates.

Second, our method provides a simpler, lower-equipment-cost alternative to SmartCut. We emphasize that our work is not intended to replace SmartCut, but rather as a complementary integration pathway with distinct advantages. SmartCut requires expensive ion implantation and specialized bonding equipment, whereas our roll-transfer process relies only on wet etching and a roll laminator, making it substantially simpler and more accessible for exploratory research and potentially pilot-scale manufacturing. Because our method is compatible with reusable bulk wafers, the overall cost can be competitive with—or even lower than—SmartCut, particularly considering the reduced capital equipment requirements. The demonstrated ability to stack more than two silicon layers vertically, again, is another critical advantage.

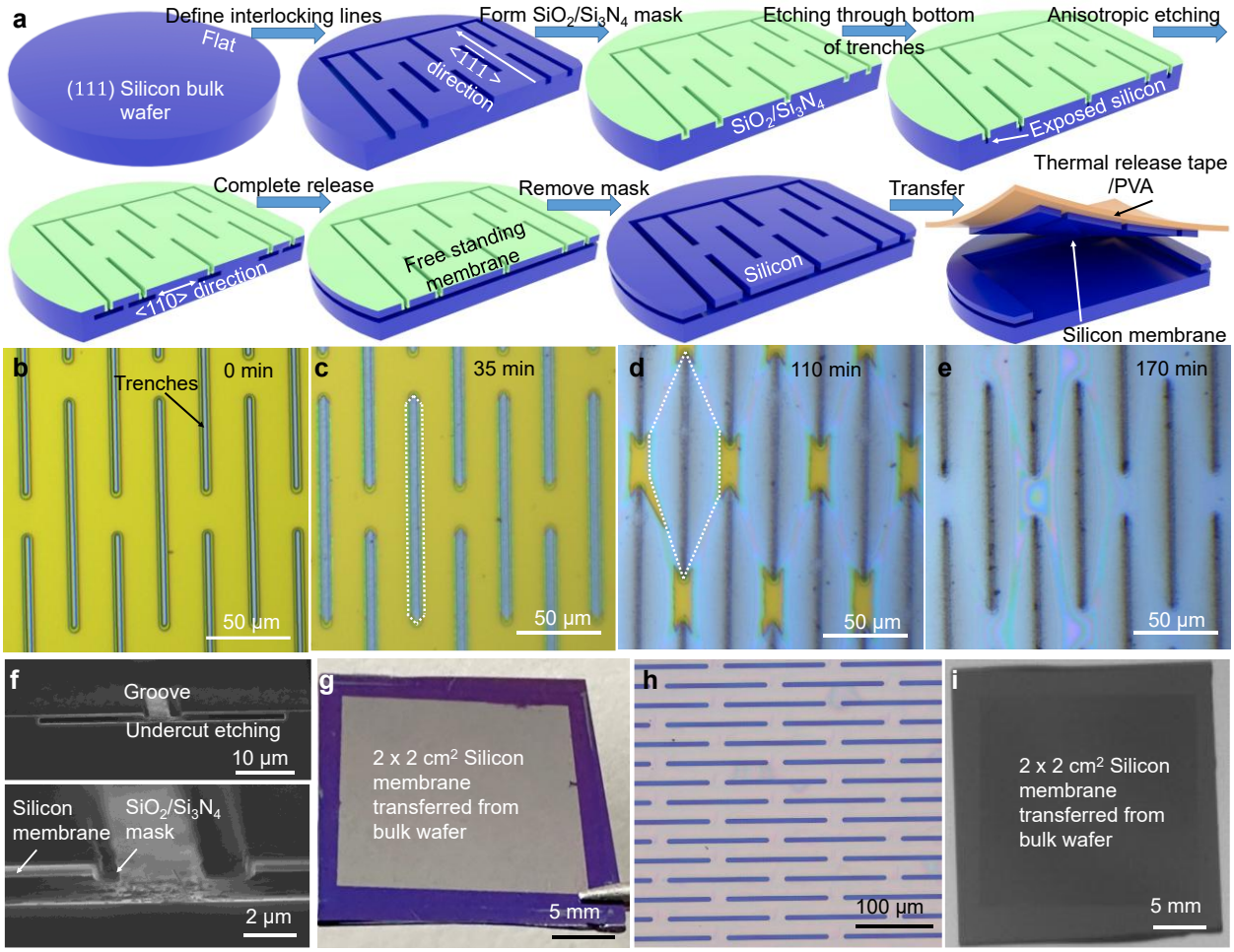
Third, the key advantage is not oxidation itself, but eliminating CMP on the transferred film. We thank the referee for pointing out the need for clearer discussion on this point. We have now explicitly clarified that a key advantage of our process lies not in defining the device layer thickness through oxidation which is standard, but more in eliminating the need for any additional CMP on the transferred silicon layer. In SmartCut combined with direct wafer bonding, CMP is required not only on the donor wafer but also on the transferred film itself after bonding to the receiving substrate to remove implantation-induced defects and restore surface roughness (as illustrated in Fig. 2 of Ref. [IEDM 2023], reproduced in Fig. S3). This CMP step introduces risks of particle contamination and within-wafer non-uniformity (WIWNU) due to pad wear and slurry flow variations. In contrast, our process defines the nanomembrane thickness through oxidation of the source wafer and transfers the film without any post-transfer polishing, thereby avoiding CMP-related damage and variability entirely. We have revised the manuscript to make this distinction explicit.



[Text redacted]

In summary, our revised manuscript now demonstrates compatibility with reusable bulk Si wafers, clarifies that our method offers a simpler, lower-equipment-cost alternative to SmartCut, and highlights the complete elimination of CMP on the transferred film, avoiding associated contamination and WIWNU risks. These additions directly address the referee's concerns and strengthen the technological positioning of our approach.

Our modification to the manuscript: We added the following statement to the revised manuscript: "*In addition to utilizing SOI wafers, freestanding silicon nanomembranes can also be derived from bulk silicon (111) wafers through anisotropic etching, providing a potentially lower-cost fabrication route (See Method and Extended Data Fig. 4)*[37]"



Extended Data Figure 4| Preparation of freestanding silicon nanomembranes from bulk silicon (111) wafers. a, Schematic illustration of the process flow. **b-e**, Optical micrographs showing the progress of <110> silicon undercut etching as a function of time from 0 min (part a), 35 min (part b), 110 min (part c), to 170 min (part d) at which point the silicon nanomembrane was fully released, for 50 μm pitch between 5 μm wide interlocking access trenches. White dashed lines serve as visual guide to mark the advancing undercut etching frontier. **f**, Cross-sectional SEM micrograph showing the suspended silicon nanomembrane following anisotropic undercut etching. **g-i**, Optical photo (part **g**), top-view optical micrograph (part **h**), and infrared image (part **i**) of a nanomembrane derived from a bulk (111) silicon wafer and transferred to a SiO_2/Si subtracted covered by spin-cast low-κ ILD.

Methods

Fabrication of freestanding single-crystalline silicon nanomembranes from bulk silicon wafers. The process is schematically illustrated in Extended Data Fig. 4a. A (111)-oriented silicon wafer (Prime grade, University Wafers) was first spin-coated with AZ5214E photoresist

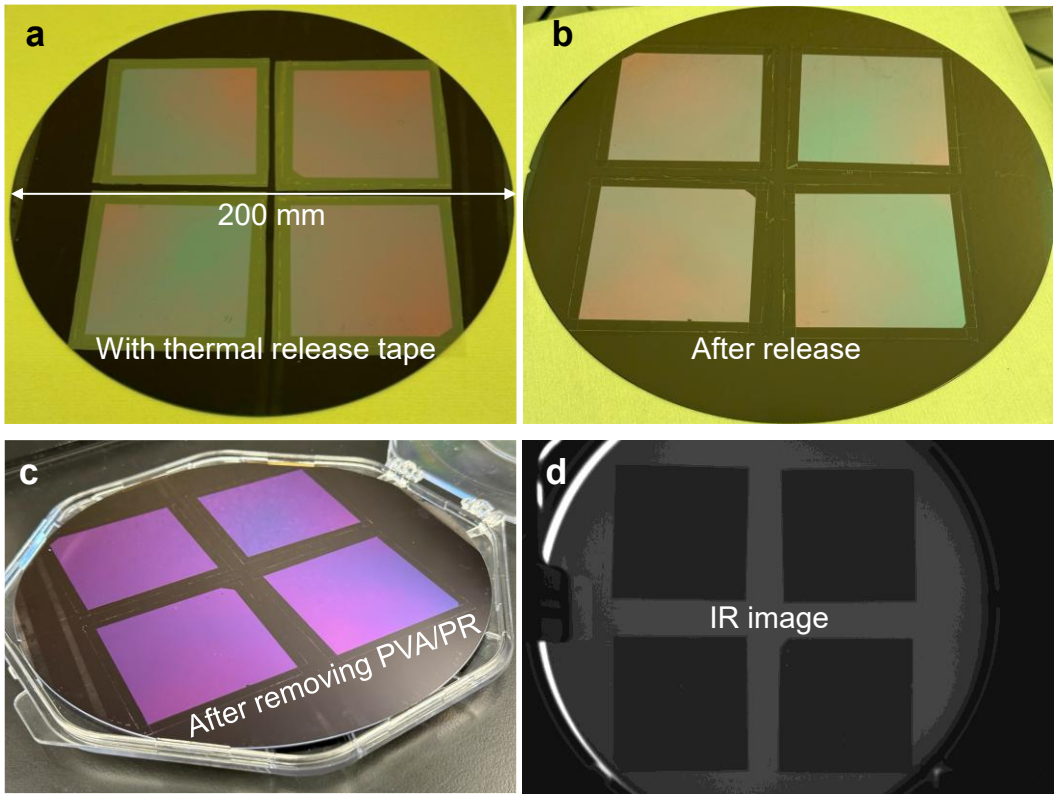
and patterned into interlocking lines, which were oriented perpendicular to the $<\bar{1}10>$ crystallographic direction. Using the photoresist as an etch mask, CF_4 RIE produced arrays of grooves, with the etching depth defining the eventual nanomembrane thickness. A 200 nm thermal oxide was then grown by dry oxidation to cover the silicon surface, followed with plasma-enhanced chemical vapor deposition of 170 nm Si_3N_4, which collectively served as the hard mask for subsequent etching. A second photolithography step defined openings within the grooves. RIE was then used to sequentially remove the oxide-nitride mask at the groove bottoms and etch the underlying silicon, producing ~700 nm deep trenches. After stripping the photoresist in hot acetone, the wafer was immersed in a preferential silicon etchant (PSE-200, Transene) at 55 ºC for anisotropic etching along the <110> and $<\bar{1}10>$ directions, with the trench sidewalls protected by the SiO_2/ Si_3N_4 bilayer mask. This undercut etching released freestanding silicon nanomembranes (Extended Data Fig. 4b-e), with the high etching anisotropy confirmed with cross-sectional SEM (Extended Data Fig. 4f). Finally, etching in BOE removed the SiO_2/Si_3N_4 mask and completed the fabrication of freestanding silicon nanomembranes from bulk silicon wafers. These membranes can then be transferred using the same roll transfer-printing process to receiving substrates for monolithic 3D stacking. Optical picture, top-view micrograph, and IR image (Extended Data Fig. 4g-i) confirms the absence of defects or voids in transferred film sitting on a silicon wafer covered with spin-cast low-κ oxide ILD.

In the revised manuscript, we cited the reference recommended by the referee for comparison and added the following statement to clarify the advantage of our process over SmartCut and wafer bonding is to avoid the polishing step: "*The silicon nanomembrane thickness is defined by oxidation of the wafer-based source material, eliminating the need for additional polishing or etching to remove surface defects or control film thickness*[41]. *This avoids risks of particle contamination and within-wafer non-uniformity associated with pad wear and slurry flow variations in chemical-mechanical polishing*[42,43]."

Comment #3: *"**Lack of detailed bond strength and defectivity study**: There are several studies of Si-SiO2 bond strength and defectivity levels necessary from CMOS technology standpoint. Typically, the defectivity is measured with a scanning acoustic microscope or IR imaging of the bond interface. An optical image can not completely reveal the minute defects that lead to process issues in advanced CMOS fabrication. Bond strength targets (DOI: 10.1088/2043-6262/1/4/04300) are also based on requirements of advanced CMOS process (CMP, stress liners, layout-based stress, etc.). Hence, annealing is performed at elevated temperatures to attain the necessary target. It is also known that Si-SiO2 interface yields good strength at 200°C (DOI: 10.1088/2043-6262/1/4/04300). The choices made in advanced Si 3D monolithic work are based on all these considerations. The claims made in the paper regarding low temperature process, are not done under these constraints."*

Our response to comment #3: We thank the referee for these constructive comments. We have performed additional experiments and expanded the manuscript to more rigorously study and quantify interface defectivity and bonding strength under conditions relevant to monolithic 3D silicon integration. Our response is organized into three parts:

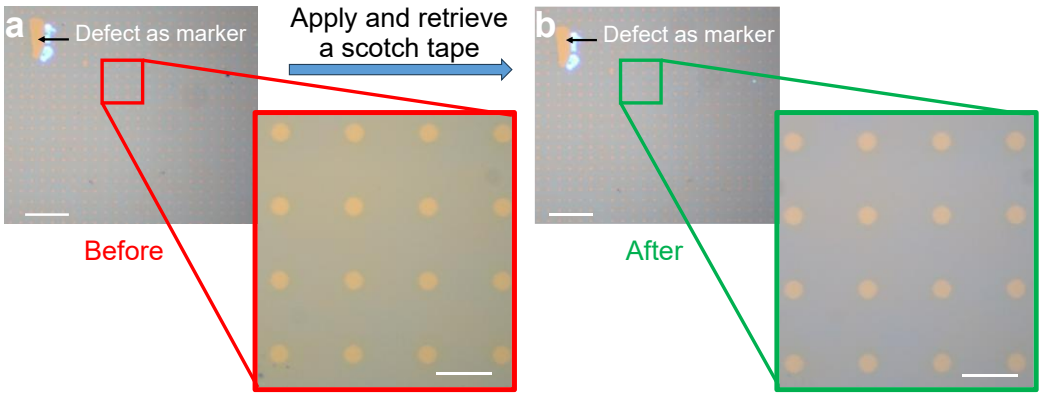
First, bond-interface defectivity: IR imaging confirms void-free bonding at wafer scale. We agree optical imaging alone is insufficient to detect minute voids at interface; therefore, we performed infrared transmission microscopy imaging on silicon membranes transferred up to covering a 200 mm wafer. The acquired IR image shows a uniform, void-free interface across the full wafer, with no detectable minute defects (Extended Data Fig. 6). This result corroborates a key advantage of our approach relative to conventional wafer bonding: the ultrathin, mechanically compliant Si nanomembrane conforms intimately to substrate topography, suppressing interfacial void formation even on surfaces with non-ideal planarity.



Extended Data Figure 6| Ultrathin, single-crystalline silicon nanomembrane transferred onto an eight-inch (200 mm-diameter) SiO_2/Si wafer substrate. a-c, Optical images showing the transferred membranes at successive stages: with the thermal release tape attached (part **a**), after release of the tape while retaining the PVA/photoresist (PR) supporting layers (part **b**), and the silicon films only after stripping the support layers (part **c**). **d**, IR image of the full wafer confirming a uniformly bonded interface with no detectable voids.

Second, bond strength: adhesion energy likely exceeds CMOS-relevant thresholds. We agree that quantitative adhesion metrics are essential for CMOS-relevant integration. However, the standard razor-blade method used in wafer bonding cannot be applied here because the transferred silicon membrane (<10 nm) is far thinner than the ~100 µm blade. Instead, we used the Scotch tape peel test as a practical adhesion benchmark. As shown in Extended Data Fig. 9, the transferred film withstands tape peeling. Based on established correlations (10.3390/mi11060605), this corresponds to an adhesion energy exceeding ~1250 mJ m^{-2}. This value is comparable to or higher than adhesion energies reported for Si-SiO_2 wafer bonding after low-temperature annealing (10.1088/2043-6262/1/4/04300) and is consistent with requirements

for downstream CMOS processes such as CMP. We have clarified this point in the revised manuscript.



Extended Data Figure 9| Strong adhesion between the transferred silicon nanomembrane and the receiving substrate as verified with scotch-tape test. a-b, Optical micrographs (scale bars: 100 μm) showing the ultrathin, single-crystalline silicon nanomembrane transferred to a SiO_2/Si substrate before (part **a**) and after (part **b**) being subjected to a scotch-tape test, with zoom-in views (scale bars: 20 μm). The structural defect was intentionally created and utilized as a marker for us to locate the exact same region for comparison.

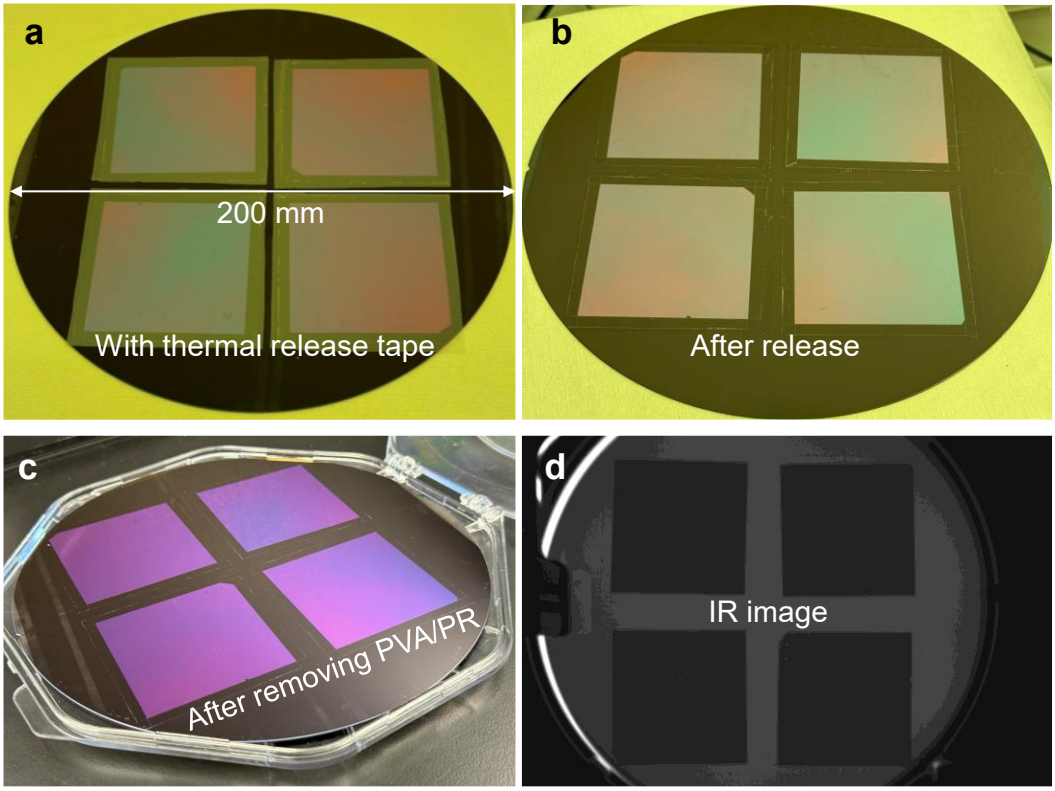
Third, thermal treatment: strengthening the Si-SiO_2 interface within a BEOL-compatible budget. We agree with the referee that thermal treatment plays a critical role in achieving strong Si-SiO_2 adhesion. In our process, annealing is performed first at 150 °C during transfer to promote initial bonding, and subsequently at 400 °C forming gas annealing during device fabrication to further strengthen the interface. The intimate, conformal contact enabled by the ultrathin membrane further enhances interfacial strength without requiring the elevated temperatures (>600 °C) used in some wafer-bonding processes. We have revised the manuscript to emphasize the combined roles of conformal contact and thermal treatment in achieving CMOS-relevant adhesion strength.

In summary, our revised manuscript now includes: IR imaging demonstrating void-free interfaces across wafer scale, adhesion-energy estimates (>1250 mJ m^{-2}) consistent with CMOS requirements, and a clarified discussion of how conformal contact and BEOL-compatible annealing together yield robust Si-SiO_2 bonding. These additions directly address the referee's concerns regarding defectivity and bond strength under CMOS-relevant constraints.

Our modification to the manuscript: We added the following statement to describe our new results on 200 mm wafer scale demonstration and IR imaging: "*Although our academic microfabrication facilities limit the maximum membrane size to four-inch wafers, repeating the transfer process four times enables coverage of an eight-inch (200 mm) wafer, with infrared (IR) imaging confirming a fully bonded interface without detectable minute voids (Extended Data Fig. 6).*" in the Method: "*The infrared (IR) images of transferred silicon nanomembranes on receiving substrates were taken by a EVG 301 megasonic cleaner and IR Inspector.*"

We added the following statement in the revised manuscript to discuss the bonding strength, especially the importance of intimate conformal contact and thermal treatments, and cited the references recommended by referee: "*The sub-10 nm thickness of the transferred silicon*

nanomembrane prevents direct measurement of bonding strength using standard razor-blade insertion. However, the transferred film adheres strongly enough to pass the Scotch-tape test (Extended Data Fig. 9)[47], suggesting that the interfacial bonding energy could exceed 1250 mJ·m^{-2}, which is consistent with the requirement for additional advanced complementary-metal-oxide-semiconductor (CMOS) processes[48]. The intimate, conformal contact with the substrate established during film transfer, together with the thermal treatment applied in the process[48], strengthens the Si–SiO$_2$ interface"



Extended Data Figure 6| Ultrathin, single-crystalline silicon nanomembrane transferred onto an eight-inch (200 mm-diameter) SiO$_2$/Si wafer substrate. a-c, Optical images showing the transferred membranes with the thermal release tape (part **a**), after release with the PVA/photoresist (PR) supporting layers (part **b**), and the silicon films only (part **c**). **d**, IR image of the wafer showing no voids.

Comment #4: *"**Low performance of junctionless transistors:** Junctionless transistors do not meet the drive current requirement of the advanced nodes, where 3D monolithic architecture is being considered as a potential scaling option. Typically, the requirements are in the order of mA/µm. The junctionless architecture suffers from high contact resistance, extension resistance, and limited threshold voltage tunability, which are key requirements in CMOS technology. The authors present 650 µA/µm for pMOS. However, this is at much higher overdrive than CMOS standard Vov=Vdd. At the standard overdrive, the drive currents are much lower. Therefore, alternative approaches are being explored: https://ieeexplore.ieee.org/document/9265026. In addition, the authors must mention 3D monolithic CMOS reports through direct wafer bonding and benchmark junctionless transistors in those reports (DOI: 10.1109/VLSIT.2018.8510705) as*

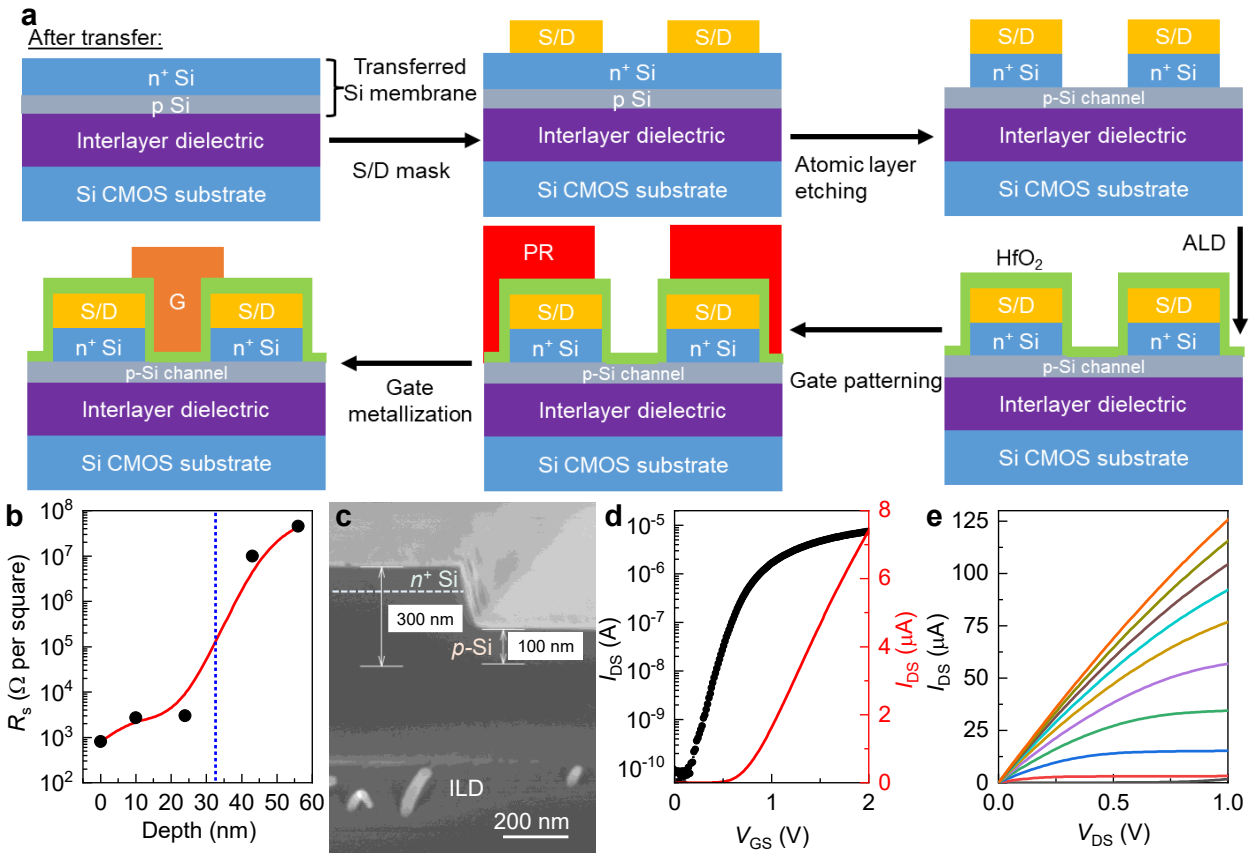
these are the most relevant for the proposed approach. Figure 2 should include the junctionless devices mentioned above in the benchmark and in the extended data figure 9. The thermal budget discussion in the paper refers to older work and is only relevant for old technology nodes. Since 3D monolithic is being considered for advanced nodes, recent work has demonstrated higher temperature tolerance (https://cea.hal.science/cea-03727300/file/VLSI2022_XGARROSv2.pdf). All of these factors need to be considered for CMOS technology."

Our response to comment #4: We thank the referee for this detailed and thoughtful comment. We have substantially expanded the manuscript to address junctionless-device performance, added benchmarking against the specific references suggested, demonstrated that our platform can also support conventional MOSFETs under a $\leqslant$400 °C thermal budget, and updated the thermal-budget discussion to reflect recent results. Our response is organized into five parts:

First, drive current density scaling of junctionless transistors: We agree that our BEOL junctionless devices, in their current form, do not meet advanced-node drive-current requirements. This is largely due to our relatively long channel length (180 nm vs. ~20 nm) and the use of a simple planar channel, reflecting the limitations of our academic fabrication facility. Even so, our junctionless devices already outperform all prior BEOL-compatible monolithic 3D demonstrations based on polysilicon, metal oxides, carbon nanotubes, and 2D semiconductors (Fig. 2h, Extended Data Fig. 14), representing a significant progress and important technology milestone. We have clarified this point in the revised manuscript.

Beyond continuous channel length scaling, a realistic path toward higher performance is to adopt 3D fin-type channels, similar to those used in advanced node MOSFETs. With a fin height of ~50 nm and a fin pitch of 30 nm (standard for 7 nm node FinFETs), the effective current density could increase by up to ~(50+8+50)/30=3.6×, pushing performance toward the mA/μm regime. We present this as a future scaling trajectory (not as a justification for current performance) in the revised manuscript.

Second, building MOSFETs on transferred silicon membranes under 400 °C to address shortcomings of junctionless transistors: We agree that junctionless devices face challenges such as high contact resistance, high extension resistance, and limited V_T tunability, despite their advantages in simplified fabrication flow requiring lower number of masks. To directly address the concern regarding the shortcomings of junctionless transistors, we have added an alternative process flow in the revised manuscript showing how our approach can also yield MOSFETs with raised very heavily doped source/drain and self-aligned, oppositely and lightly doped recessed channel regions. This process still avoids post-transfer doping and activation, remaining within a 400 °C thermal budget. Preliminary results are included in Extended Data Fig. 11 with fabrication details provided in Methods. This establishes that our monolithic 3D integration method is not limited to junctionless devices and can support standard CMOS transistor architectures.



Extended Data Figure 11| MOSFETs fabricated on transferred silicon nanomembrane under a BEOL-compatible thermal budget. **a,** Schematic illustration of the fabrication flow. **b**, Evolution of silicon sheet resistance with ALE etching depth, measured by four-point probe. The blue dashed line serves as a visual guide to mark the approximate boundary position between the heavily doped surface layer and the oppositely and lightly doped channel. **c**, Cross-section SEM micrograph showing the heavily *n*-doped raised source-drain regions and light *p*-doped recessed channel defined by etching. **d**, Transfer characteristics of a completed MOSFET, plotted in both linear (red, right axis) and logarithmic (black, left axis) scales. Applied V_{DS}=0.1 V. **e**, Corresponding output characteristics. The device L_{ch}=3 μm and W=2 μm. The V_{GS} descends from 4.5 V at a step of 0.5 V from top to bottom.

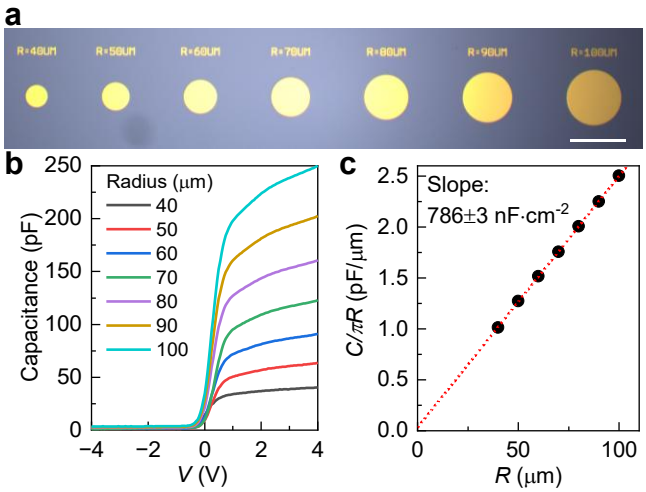
In 10.1109/VLSITechnology18217.2020.9265026, excimer laser annealing was employed for dopant activation. Compared to conventional annealing, excimer lasers confine high-temperature region and thereby reduce the risk of damage to bottom-tier devices and interconnects. Nevertheless, thermal spreading remains a concern for monolithically integrated 3D circuits, particularly when very thin interlayer dielectrics are used to achieve high vertical interconnect density. For example, Fig. 12 of [10.1109/TED.2024.3460763] reports that the oxide interlayer dielectric must be at least 300 nm thick to ensure the substrate temperature stays below 400 °C during a 13 ns laser annealing of the top silicon layer (Fig.S4). In contrast, our approach does not face this limitation and enables scaling of the interlayer dielectric thickness to well below 100 nm (see Fig. 1f and Fig. 2c), as demonstrated experimentally.

[Figure redacted]

[Text redacted]

Third, drive voltage V_{DD} as limited by interfacial oxide and EOT scaling: Our devices currently require higher V_{DD} due to the large equivalent oxide thickness (EOT). For comparison, TSMC's 3 nm FinFETs operate at ~0.8 V with an EOT of ~0.7 nm, while our devices exhibit an EOT of 4.4 nm measure in experiment (Extended Data Fig. 15). The present EOT in our devices is limited by the ~2 nm interfacial SiO_2 observed in STEM (Fig. 1f inset). This ~6× larger EOT directly translates into the need for proportionally higher gate overdrive, explaining our present 2–4 V operation. A discussion of this limitation and outlook has been added.

Industry has developed several well-established approaches to minimize the contribution from interfacial oxide to EOT: GlobalFoundries has described native oxide removal in de-oxygenated HF followed by immediate high-k deposition; IBM has discussed interfacial layer thinning via scavenging reactions (10.3390/ma5030478); and Applied Materials has reported plasma nitridation to increase the effective dielectric constant of the interlayer oxide (10.1149/1.2911492). These approaches require specialized equipment but are fully compatible with our process and would significantly reduce EOT—and thus V_{DD}—in future development. This is not an intrinsic limitation of our method.



Extended Data Figure 15| Gate capacitance of fabricated junctionless transistors. a, Optical image of an array of circular *n*-doped silicon nanomembrane/HfO_2/Cr/Au capacitors with

their radii (R) varied from 40 to 200 μm. Scale bar: 100 μm. **b**, Measured capacitance-voltage characteristics of the capacitors shown in part **a**. **c**, Capacitance (C) to πR ratio as a function of R. The red dashed line represents a linear fit to the data (black dots) to extract the capacitance per area of the native oxide plus 10 nm HfO_2 stack from its slope. The y-axis intercept gives the capacitance caused by the fringing field from the edge of electrodes.

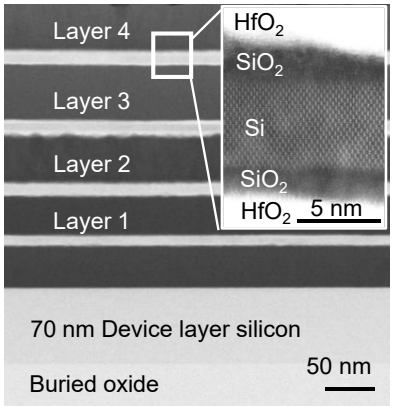


Figure 1. Wafer-scale, monolithic 3D integration of single-crystalline silicon nanomembranes. **f**, STEM image showing the cross-section of a stacked 5 tiers of single-crystalline silicon on a monolithic substrate. Scale bar: 50 nm. Inset: Zoom-in view of a single tier, scale bar: 5 nm.

Fourth, benchmark against direct wafer bonding approach: We agree that direct wafer bonding is an important reference point. Following the referee's suggestion, we have added benchmarking in Fig. 2h and Extended Data Fig. 14 against the junctionless devices reported in Ref. [10.1109/VLSIT.2018.8510705], the exact reference the reviewer requested. Our devices compare favorably in terms of normalized current density and transconductance. We also note that, while our approach is at an earlier stage of development, it has already demonstrated integration of three BEOL-compatible transistor layers, compared to only one top layer in the wafer bonding approach in the previous works. Additionally, our entire process remains below 400 °C, whereas wafer-bonding-based demonstrations generally require >500 °C.

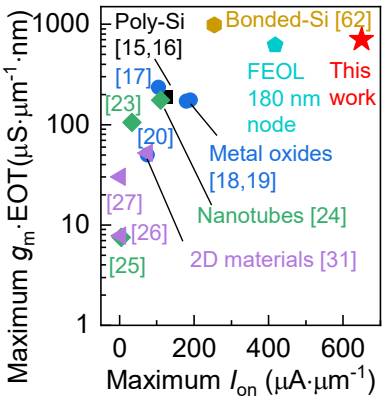
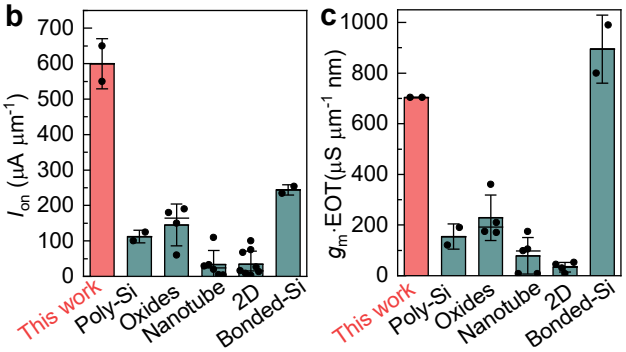


Figure 2. BEOL-compatible junctionless transistors built on transferred single-crystalline silicon nanomembranes. **h**, Plot of g_m·EOT *versus* I_{on} to compare with best BEOL-compatible transistors reported in literature, bonded 40 nm-node monolithic 3D silicon

transistors, and 180 nm-node FEOL silicon MOSFET whose performance characteristics are extracted from a standard-cell library.

a

Materials	CMOS	Mobility ($cm^2 V^{-1}s^{-1}$)	Maximum I_{on} ($\mu A \cdot \mu m^{-1}$)	Normalized Maximum $g_m \cdot EOT$ ($\mu S \cdot \mu m^{-1} \cdot nm$)	L_g (nm)	EOT (nm)	Device type	Thermal budget	Ref
Single-crystalline silicon	Yes	50±10 (*n*-FET, up to 85) 60±20 (*p*-FET, up to 87)	650 (*p*-FET) 550 (*n*-FET)	704 (*p*-FET) 704 (*n*-FET)	180	4.4	Planar junctionless FET	≤400 °C with transfer	This work
Laser annealed crystalline silicon island	Yes	N.R.	125 (*n*-FET) 100 (*p*-FET) Electrical *W*	190 (*n*-FET) 120 (*p*-FET)	150	~1.25	FinFET	~2200 °C at Si surface; ILD>400 nm	[15,16]
C-axis-aligned IGZO	*n*-FET only	5–10	60–150	210–360	72	6	Planar	<400 °C	[17]
Polycrystalline IGZO	*n*-FET only	N.R.	180	170	60	~2.5	Double gate-FET	<400 °C	[18]
In_2O_3	*n*-FET only	15	190	175	100	~7	Double gate -FET	Room temperature but device only stable up to 50 °C	[19]
In_2O_3/organic	Yes	4±2 (*n*-FET) 3±2 (*p*-FET)	76 (*n*-FET)	52	100	~7	Planar	only stable up to 50 °C	[20]
Aligned nanotubes	*p*-FET only	N.R.	20	N.R.	1000	N.R.	Planar	<200 °C with transfer	[22]
Nanotube random network	Yes	N.R.	33 (*p*-FET) 29 (*n*-FET)	105 (*p*-FET) 98 (*n*-FET)	130	3–4	Planar bottom gate	<425 °C with solution deposition	[23,73]
Nanotube random network	Yes	N.R.	110	175	100	3–4	Planar bottom gate	<415 °C with solution deposition	[24,73]
Nanotube random network	Yes	10–25	4.5 (*p*-FET) 3.6 (*n*-FET)	7.5 (*p*-FET) 7.5 (*n*-FET)	2000	~3	Planar bottom gate	300 °C with transfer	[25]
MoS_2	*n*-FET only	35.9	2.9	7.8	200	8.7	Planar top gate	Growth at 300 °C	[26]
Bilayer MoS_2	*n*-FET only	10.4	2.8	30	5000	16	Planar top gate	120 °C with transfer	[27]
MoS_2 and WSe_2	Yes	N.R.	99.8 (*n*-FET) 12.9 (*p*-FET)	N.R.	45	6	Planar bottom gate	180 °C with transfer	[28]
MoS_2	*n*-FET only	42	8	N.R.	2000	5	Planar bottom gate	180 °C with transfer	[29]
WSe_2	Yes	6.9 (*n*-FET) 7 (*p*-FET)	26 (*n*-FET) 16 (*p*-FET)	N.R.	300	6	Planar bottom gate	180 °C with transfer	[30]
MoS_2 and WSe_2	Yes	56 (*n*-FET) 51 (*p*-FET)	67 (*n*-FET) 75 (*p*-FET)	52 (*n*-FET) 44 (*p*-FET)	400	1.56	Planar top gate	Growth at 385 °C	[31]
Single-crystalline silicon	Yes	N.R.	254 (*p*-FET) 234 (*n*-FET)	990 (*p*-FET) 800 (*n*-FET)	40	1.66	Planar top gate	525 °C	[62]



Extended Data Figure 14| Performance comparison between monolithic 3D transistors. **a,** Table listing the comparison of key device characteristics. **b-c**, Comparison of width-

normalized I_{on} (part **b**) and g_m·EOT (part **c**). GAA: gate-all-around; IGZO: indium gallium zinc oxide; ILD: interlayer dielectric.; N.R.: not reported.

Fifth, updated thermal-budget discussion for advanced nodes: We appreciate the referee's point regarding more recent studies showing improved thermal tolerance. We have added the recommended reference (DOI: 10.1109/VLSITechnology18217.2020.9265075), which reports tolerance up to ~500 °C for 28 nm technology. We have clarified in the revised manuscript that although such results are encouraging, it remains prudent from an engineering perspective to target substantially below 500 °C to preserve margin for yield and long-term reliability, against Cu diffusion due to additional barrier thinning and degradation of even lower-k (more porous) interlayer dielectrics in further scaled interconnects in future nodes. Our $\leqslant$ 400 °C target therefore remains aligned with conservative BEOL-integration requirements for advanced nodes.

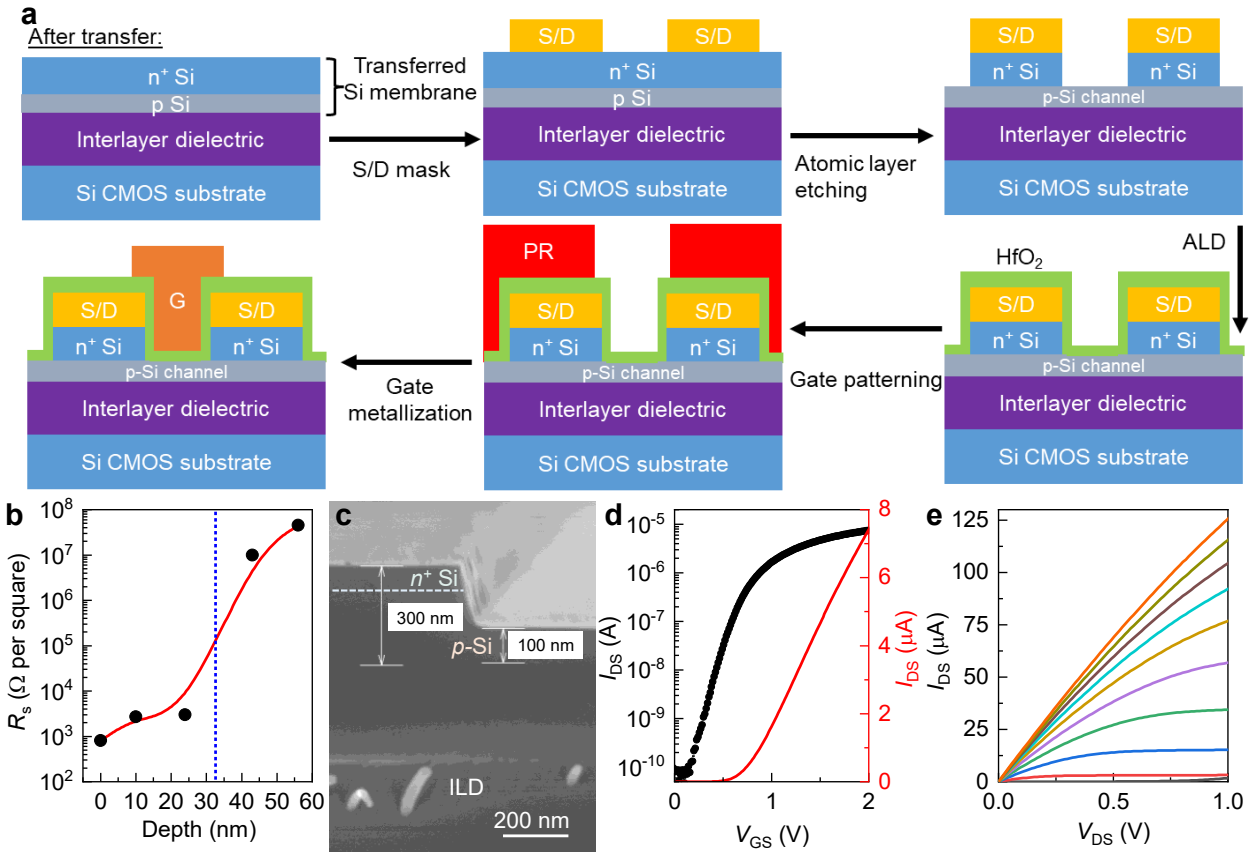
Our modification to the manuscript: In the revised manuscript, we added the following statement to acknowledge the lower drive current than state-of-the-art FEOL MOSFETs and suggested approaches to improve: "*Their output current density remains below that of FEOL MOSFETs in more advanced technology nodes, which surpass 1 mA·μm^{-1}, as limited mainly by the planar channel structure. Future iterations that transition from planar channels to vertical fins are expected to significantly narrow this performance gap.*"

In the revised manuscript, we added the following statement to describe the option of fabricating MOSFETs instead of junctionless transistors with transferred Si nanomembranes: "*Meanwhile, MOSFETs incorporating raised, heavily doped source/drain regions and oppositely doped, recessed channels can also be fabricated from the transferred silicon nanomembranes within a <400 °C thermal budget (See Methods and Extended Data Fig. 11). Compared to junctionless transistors, such MOSFETs alleviate some limitations, including high contact and extension resistance as well as restricted threshold tunability, at the expense of a more complex fabrication flow.*"

Methods

Fabrication of silicon MOSFETs with raised source-drain and recessed channel: The process is schematically illustrated in Extended Data Fig. 11a. A SOI wafer with a 300 nm lightly *p*-doped (~10^{15} cm^{-3}) device layer was first blanket-coated with SOD (Filmtronics P505). Rapid thermal annealing was performed with a temperature ramp rate of 50 °C s^{-1} to 950 °C, followed by rapid cooling under flowing argon (1000 s.c.c.m.), producing a heavily *n*-doped surface layer ~35 nm deep (Extended Data Fig. 11b). After layer transfer, a AZ5214E photoresist film was spin-cast at 3000 rpm for 30 s. Alignment marks and source–drain patterns were defined in the resist, aligned to the mesh pattern of the transferred nanomembranes, and processed through development, sputtering of 0.4 nm Cr/60 nm Au, and lift-off in hot acetone. The resulting source–drain electrodes served as a hard mask for atomic-layer etching (ALE) of silicon using an Oxford mixed inductively coupled plasma (ICP)-RIE system. Each ALE cycle consisted of a purge step (RIE power was 0 W) and an etch step (RIE power was 30 W), with ICP power fixed at 750 W and gas flows of 40 s.c.c.m. SF_6 and 100 s.c.c.m. Ar. Each cycle removed ~1 nm of silicon with low surface roughness with the ALE smoothing effect[72]. After ~200 cycles, a 100 nm thick lightly *p*-doped recessed channel was formed, self-aligned to the raised, heavily *n*-doped source-drain and metal contacts (Extended Data Fig. 11c). A second photolithography step patterned etch masks to protect the channel regions, followed by RIE to

remove unprotected silicon, thereby defining device width and achieving device isolation. The resist was stripped in acetone, and an oxygen plasma clean was performed before depositing a 10 nm HfO_2 gate dielectric by ALD. A third photolithography step defined the gate electrodes, followed by sputtering of a 4 nm Cr/40 nm Au stack and lift-off in acetone. Finally, a fourth photolithography step opened contact windows for probing pads, and dry etching of the HfO_2 exposed the underlying contacts. Resist removal in acetone completed the transistor fabrication, all within a thermal budget $\leqslant 400\,^{\circ}C$. The resulting devices exhibited effective mobilities exceeding 90 $cm^2V^{-1}s^{-1}$, as extracted from their transfer and output characteristics (Extended Data Fig. 11d and 11e).



Extended Data Figure 11| MOSFETs fabricated on transferred silicon nanomembrane under a BEOL-compatible thermal budget. **a,** Schematic illustration of the fabrication flow. **b**, Evolution of silicon sheet resistance with ALE etching depth, measured by four-point probe. The blue dashed line serves as a visual guide to mark the approximate boundary position between the heavily doped surface layer and the oppositely and lightly doped channel. **c**, Cross-section SEM micrograph showing the heavily *n*-doped raised source-drain regions and light *p*-doped recessed channel defined by etching. **d**, Transfer characteristics of a completed MOSFET, plotted in both linear (red, right axis) and logarithmic (black, left axis) scales. Applied V_{DS}=0.1 V. **e**, Corresponding output characteristics. The device L_{ch}=3 μm and W=2 μm. The V_{GS} descends from 4.5 V at a step of 0.5 V from top to bottom.

In the revised manuscript, we added the following statement explaining the large V_{DD}: "*Our BEOL-compatible junctionless transistors employ a gate dielectric stack comprising ~1.5–2 nm native interfacial SiO_2 and 10 nm HfO_2, corresponding to an EOT of 4.4 nm (Extended Data Fig.*

15). This EOT is roughly twice that of 180 nm-node MOSFETs (EOT=2.2–3 nm)[64,65], ~2.7 times larger than 40 nm-node bonded silicon transistors[62], and more than six times greater than the ~0.7 nm EOT of state-of-the-art 3 nm technology node. Such a large EOT necessitates a higher gate overdrive voltage. Reducing the interfacial native oxide through in situ removal prior to gate oxide growth, scavenging reactions, or nitridation treatments should lower the EOT[66,67], hereby reducing the required drive voltage V_{DD} and further improving device performance and scalability."

In the revised manuscript, we added benchmarking in Fig. 2h and Extended Data Fig. 14 against the devices reported in DOI:10.1109/VLSIT.2018.8510705: "*We further compare our best-performing p-FETs with commercial 180 nm-node front-end-of-line (FEOL) silicon MOSFETs which have similar device dimensions and monolithic 3D silicon transistors formed by wafer bonding with a higher thermal budget of 525 °C[62].*"

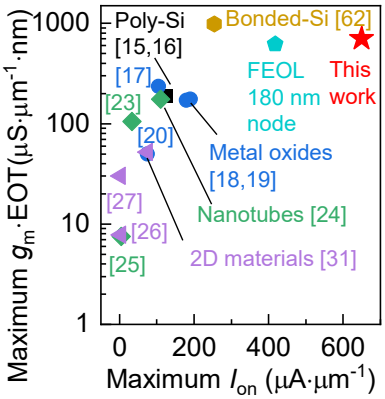
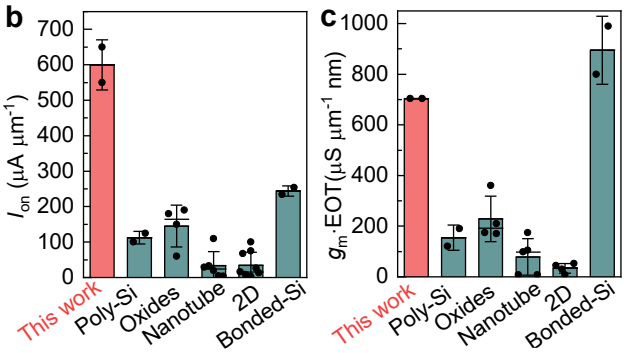


Figure 2. BEOL-compatible junctionless transistors built on transferred single-crystalline silicon nanomembranes. h, Plot of g_m·EOT *versus* I_{on} to compare with best BEOL-compatible transistors reported in literature, bonded 40 nm-node monolithic 3D silicon transistors, and 180 nm-node FEOL silicon MOSFET whose performance characteristics are extracted from a standard-cell library.

a

Materials	CMOS	Mobility ($cm^2 V^{-1}s^{-1}$)	Maximum I_{on} ($\mu A \cdot \mu m^{-1}$)	Normalized Maximum g_m·EOT ($\mu S \cdot \mu m^{-1} \cdot nm$)	L_g (nm)	EOT (nm)	Device type	Thermal budget	Ref
Single-crystalline silicon	Yes	50±10 (*n*-FET, up to 85) 60±20 (*p*-FET, up to 87)	650 (*p*-FET) 550 (*n*-FET)	704 (*p*-FET) 704 (*n*-FET)	180	4.4	Planar junctionless FET	≤400 °C with transfer	This work
Laser annealed crystalline silicon island	Yes	N.R.	125 (*n*-FET) 100 (*p*-FET) Electrical *W*	190 (*n*-FET) 120 (*p*-FET)	150	~1.25	FinFET	~2200 °C at Si surface; ILD>400 nm	[15,16]
C-axis-aligned IGZO	*n*-FET only	5–10	60–150	210–360	72	6	Planar	<400 °C	[17]
Polycrystalline IGZO	*n*-FET only	N.R.	180	170	60	~2.5	Double gate-FET	<400 °C	[18]
In_2O_3	*n*-FET only	15	190	175	100	~7	Double gate -FET	Room temperature but device only stable up to 50 °C	[19]
In_2O_3/organic	Yes	4±2 (*n*-FET) 3±2 (*p*-FET)	76 (*n*-FET)	52	100	~7	Planar	only stable up to 50 °C	[20]

Aligned nanotubes	*p*-FET only	N.R.	20	N.R.	1000	N.R.	Planar	<200 °C with transfer	[22]
Nanotube random network	Yes	N.R.	33 (*p*-FET) 29 (*n*-FET)	105 (*p*-FET) 98 (*n*-FET)	130	3–4	Planar bottom gate	<425 °C with solution deposition	[23,73]
Nanotube random network	Yes	N.R.	110	175	100	3–4	Planar bottom gate	<415 °C with solution deposition	[24,73]
Nanotube random network	Yes	10–25	4.5 (*p*-FET) 3.6 (*n*-FET)	7.5 (*p*-FET) 7.5 (*n*-FET)	2000	~3	Planar bottom gate	300 °C with transfer	[25]
MoS_2	*n*-FET only	35.9	2.9	7.8	200	8.7	Planar top gate	Growth at 300 °C	[26]
Bilayer MoS_2	*n*-FET only	10.4	2.8	30	5000	16	Planar top gate	120 °C with transfer	[27]
MoS_2 and WSe_2	Yes	N.R.	99.8 (*n*-FET) 12.9 (*p*-FET)	N.R.	45	6	Planar bottom gate	180 °C with transfer	[28]
MoS_2	*n*-FET only	42	8	N.R.	2000	5	Planar bottom gate	180 °C with transfer	[29]
WSe_2	Yes	6.9 (*n*-FET) 7 (*p*-FET)	26 (*n*-FET) 16 (*p*-FET)	N.R.	300	6	Planar bottom gate	180 °C with transfer	[30]
MoS_2 and WSe_2	Yes	56 (*n*-FET) 51 (*p*-FET)	67 (*n*-FET) 75 (*p*-FET)	52 (*n*-FET) 44 (*p*-FET)	400	1.56	Planar top gate	Growth at 385 °C	[31]
Single-crystalline silicon	Yes	N.R.	254 (*p*-FET) 234 (*n*-FET)	990 (*p*-FET) 800 (*n*-FET)	40	1.66	Planar top gate	525 °C	[62]



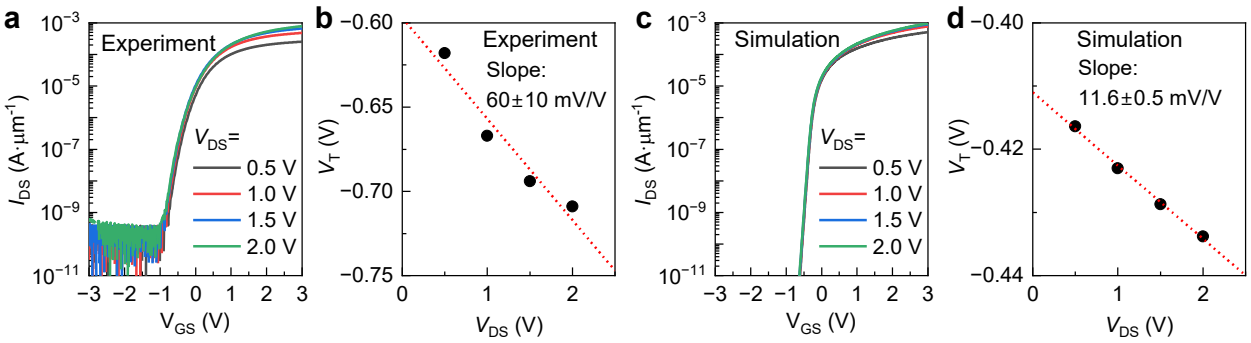
Extended Data Figure 14| Performance comparison between monolithic 3D transistors. **a,** Table listing the comparison of key device characteristics. **b-c**, Comparison of width-normalized I_{on} (part **b**) and g_m·EOT (part **c**). GAA: gate-all-around; IGZO: indium gallium zinc oxide; ILD: interlayer dielectric.; N.R.: not reported.

In the revised manuscript, we cited the references recommended by the referee and added the following discussion regarding the thermal budget limit for monolithic 3D integration: "*However, a key challenge of implementing monolithic 3D integration is to form top-tier semiconductors and transistors within a limited thermal budget ideally ≤ 400 °C to preserve bottom-tier devices and interconnects from degradation*[3]. *Previous studies have shown that annealing at 475 °C for 1 hour increases via resistance by more than fourfold in the 250 nm technology node*[12]*; while annealing at 550 °C for 2 hours results in over 90% yield loss in the 28 nm node*[13]. *As interconnect dimensions and barrier layer thickness continue to scale, the allowable thermal budget is expected to become even more restrictive in future technology nodes.*"

Comment #5: *"The authors should present drain induced barrier lowering (DIBL) and subthreshold slope at high Vd in Figure 2. This is a necessary characterization for short channel devices for digital applications."*

Our response to comment #5: We thank the referee for this valuable suggestion. We agree that DIBL and the subthreshold slope under high drain bias are critical metrics for assessing short-channel devices intended for digital applications. In response, we have therefore added transfer characteristics measured across V_{DS} values for our short-channel junctionless transistors. These measurements show that the sharp subthreshold slope remains essentially unchanged at high drain bias, and the DIBL remains small at 60±10 mV/V, which is only modestly higher than the 11.6 ±0.5 mV/V predicted by TCAD simulation that includes electrostatics alone. These results confirm that the devices maintain strong electrostatic control even under high drain bias. The new data and discussion are now included in Extended Data Fig. 13 and referenced in the main text.

Our modification to the manuscript: We added the following statement in the revised manuscript: "*These short-channel junctionless transistors exhibited a small drain induced barrier lowering (DIBL) of 60±10 mV/V, which is close to the value predicted by TCAD simulations, with the remaining discrepancy likely arising from silicon-HfO₂ interface traps (Extended Data Fig. 13).*"



Extended Data Figure 13| DIBL in scaled junctionless transistors. **a,** Experimentally measured transfer characteristics of a junctionless transistor with L_{ch} of ~ 200 nm, under applied V_{DS} of 0.5 V (black), 1.0 V (red), 1.5 V (blue), and 2.0 V (green). **b**, Device V_T as a function of applied V_{DS}. The dashed line represents a linear fit to the data used to extract DIBL from the slope. **c,** Simulated transfer characteristics of a junctionless transistor with L_{ch} of 180 nm, under the same set of applied V_{DS} values. **d**, Simulated V_T as a function of applied V_{DS}. The dashed line represents the linear fit used to extract DIBL.

Comment #6: *"**3D circuits:** The discussion surrounding 3D circuits has a few statements regarding the reduction in area compared to 2D counterparts (lines 22 and 26, page 8). The actual footprint reduction is less than what is claimed. This is understood in 3D circuits, as the overhead due to 3D via, metal placement does not provide 3 times reduction for every cell (see: DOI:10.1109/ISLPED.2017.8009189). The authors should mention it appropriately."*

Our response to comment #6: We thank the referee for raising this important point. We agree that our original phrasing may have implied an idealized footprint reduction that does not fully account for practical layout overheads in monolithic 3D circuits. As noted in the cited reference (DOI: 10.1109/ISLPED.2017.8009189), the achievable area reduction in bilayer 3D logic is ~44% rather than the ideal 50%, due to contributions from high-density 3D vias, routing congestion, and metal-layer placement. In the revised manuscript, we have clarified this

distinction by explicitly separating the theoretical maximum reduction from the practical reduction achievable after layout optimization. We now state that while monolithic 3D integration can, in principle, reduce the footprint by up to a factor of two (or three for three-tier SRAM), and then emphasize the actual realized savings are slightly smaller but still represent a substantial improvement over 2D layouts.

Our modification to the manuscript: We added the reference recommended by the referee and modified the statement in revised manuscript as: "*By employing two p-FETs and two n-FETs on different tiers, 3D NOR (Fig. 4e) and NAND gates (Fig. 4f) were also demonstrated. Compared to 2D circuits constructed using laterally adjacent MOSFETs, monolithic 3D gates achieve the same functionality but with up to two times higher device integration density (The effective area reduction may be slightly less than the ideal 50%, for example, ~44% after layout optimization[71], owing to overhead from high-density 3D vias, routing congestion, and metal placement).*" and "*The monolithic approach provides continuous scalability in the vertical dimension. Folding a six-transistor SRAM cell into three tiers of silicon junctionless transistors further reduces its circuit footprint by up to three times with balanced count of devices in each layer (Fig. 4g and 4h). As with bilayer circuits, the practical area reduction is slightly less than the ideal value due to routing and via-placement overheads.*"

Comment #7: *"Although the layer transfer process is novel, the transistor and circuit demonstrations are complex and exhaustive (considering it was performed in an academic lab), the work misses important technological aspects of relevance for Si CMOS 3D monolithic integration and circuits. Therefore, it is not impactful to advance 3D monolithic CMOS technology. Several advanced demonstrations already exist (DOI: 10.1109/IEDM45741.2023.10413864), and there are reports of several generations of improvements in device architecture (including junctionless transistors), layer transfer, and bonding methods. Highlighting demonstrations that bring out the novelty of the approach- ability to conformally transfer on topography and then demonstrate relevant devices that require such topography (as shown in: https://www.nature.com/articles/s41467-021-27208-5#Sec10) would be more appropriate than the current proposal of 3D monolithic CMOS."*

Our response to comment #7: We thank the referee for recognizing our work as "*novel, complex, and exhaustive*," and we appreciate the thoughtful perspective on the broader technological context for a broad *Nature* readership. We respectfully disagree with the conclusion that the work is not impactful for advancing monolithic 3D CMOS, and we have substantially revised the manuscript to clarify the technological relevance and unique contributions of our approach, in line with the referee's suggestions.

1. Addressing the key technological concerns raised by the referee. Across the revised manuscript, we now explicitly address the aspects the referee identified as essential for CMOS relevance:

- **Area efficiency:** We quantify and minimize the release-hole area overhead to a practically negligible level (<0.01%), supported by new experiments and analysis.
- **Cost and manufacturability:** We demonstrate with new experimental results that freestanding silicon nanomembranes can be derived from reusable bulk wafers, eliminating reliance on SOI and enabling a more cost-effective fabrication pathway.

- **Bonding quality:** We provide 200 mm-scale IR imaging and adhesion tests confirming strong bonding strength and low interface defectivity, enabled by the conformal contact of mechanically compliant ultrathin silicon membranes.
- **Device performance and MOSFET option:** We outline clear pathways to improve device drive current and reduce V_{DD}, and we experimentally validate that our approach can also support stacked MOSFETs within a $\leqslant$400 °C thermal budget, addressing limitations of junctionless devices.
- **Benchmarking:** We expand our benchmarking to include the prior monolithic 3D CMOS demonstrations recommended by the referee, enabling a more complete and transparent comparison.

2. Complementary to wafer-bonding-based monolithic 3D CMOS, providing unique and technologically important capabilities. We agree with the reviewer that the ability to conformally transfer ultrathin single-crystalline silicon onto non-planar surfaces is a central novelty. This capability is not available in wafer-bonding-based layer transfer, which requires sub-nanometer planarity and CMP. It is central to our ability to achieve three or more vertically stacked device layers with high yield, while tolerating ILD surface roughness and non-idealities, avoiding wafer warpage, and maintaining a low thermal budget. We have emphasized this point more clearly in the revised manuscript, as it represents one of the core technological contributions of our work to monolithic 3D integration.

In addition, our method achieves this multilayer stacking without requiring specialized ion implantation/wafer-bonding tools, CMP steps, solid-state epitaxy, or nanosecond laser annealing, enabling a cost-effective and equipment-light pathway that is attractive for exploratory research and pilot-scale manufacturing. We view this as a complementary integration route alongside wafer-bonding-based approach developed by *Leti*, particularly in applications where topography tolerance, number of vertical layer scalability, and low thermal budget are essential.

3. Relevance to industry and future applications. This work is supported directly by industry members of the NSF Industry-University Research Partnerships (IUCRC) Center for Advanced Semiconductor Chips with Accelerated Performance (ASAP), including IBM, TSMC, Intel, AMD, and MIT Lincoln Laboratory. Their selection of this project to be supported with their membership contributions and technological engagement reflect their views that this technology offers a promising and complementary route for both commercial and defense-oriented monolithic 3D integration. We also appreciate the referee's suggestion regarding applications in deformable electronic and optoelectronic systems, which we plan to explore in future work.

Our modification to the manuscript: Please see our responses to other comments.

Referee #4:

Summary comments: *"The authors reported wafer-scale roll transfer of ultrathin single-crystalline silicon nanomembranes to achieve monolithic 3D integration of complementary junctionless transistors under BEOL-compatible processing temperature (<400 °C). They further demonstrated high-performance p- and n-type junctionless FETs, inverters, NAND/NOR logic gates, and SRAM cells across three tiers. The manuscript is suitable in Nature, but there are several points that need to be addressed."*

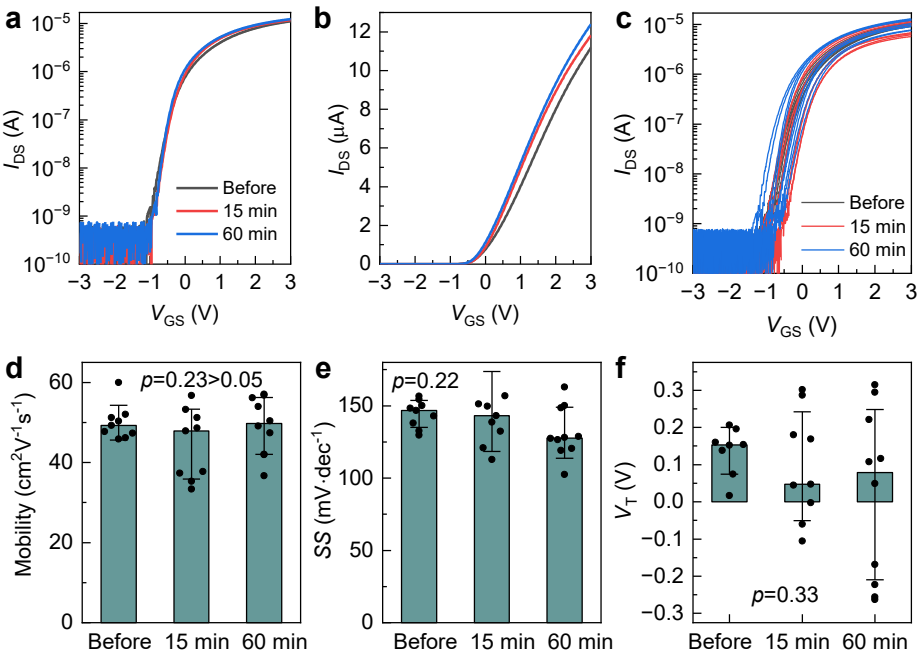
Our response: We sincerely thank the referee for the positive assessment of our work and for considering it suitable for publication in *Nature*, pending satisfactory revision. We appreciate the constructive feedback and have carefully addressed all points raised.

Comment #1: *"The authors demonstrate device performance uniformity; however, during the transfer process, significant residues can remain. When exposed to thermal treatments, these residues may induce additional variability in device performance. It would strengthen the manuscript to include thermal stability tests at 400 °C as a function of annealing time to quantify how transfer-induced residues affect device performance."*

Our response to comment #1: We thank the referee for this insightful comment. We agree that any transfer-induced residues could, in principle, introduce additional variability during subsequent thermal treatments, and that their impact should be explicitly evaluated. We employ a bilayer of photoresist and electronic-grade polyvinyl alcohol to protect the silicon surface during transfer. While this approach is highly effective in minimizing surface contamination, we performed dedicated thermal-stability tests to directly assess this concern.

In the revised manuscript, we now include annealing experiments at 400 °C—the BEOL-relevant thermal limit—for durations up to 60 min. As shown in Extended Data Fig. 21, the transfer characteristics and extracted device parameters (mobility, subthreshold swing, and threshold voltage) exhibit no statistically significant changes after annealing. One-way ANOVA yields p-values of 0.23, 0.22, and 0.33, confirming that any potential transfer-induced residues have a negligible effect on device performance under these thermal conditions. We have incorporated these new data and the corresponding discussion into the revised manuscript.

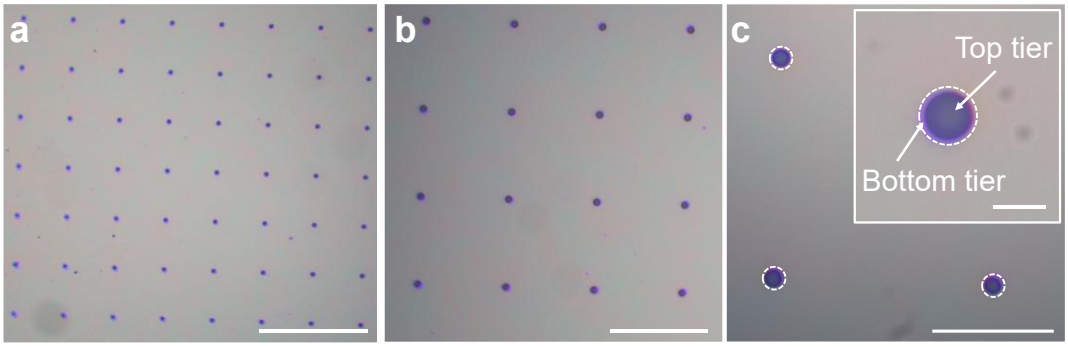
Our modification to the manuscript: We added the following statement to the revised manuscript: "*These transistors built on transferred silicon nanomembranes also exhibit good thermal stability. Devices annealed at 400 °C for 15 or 60 minutes show no statistically significant changes in their characteristics (Extended Data Fig. 21), with one-way analysis of variance (ANOVA) yielding p-values much greater than 0.05.*"



Extended Data Figure 21| Thermal stability of BEOL-compatible junctionless transistors built on transferred silicon nanomembranes. **a-b,** Transfer characteristics of a representative transistor measured before (black) and after forming-gas annealing at 400 ºC for 15 min (red) and 60 min (blue) plotted on logarithmic (part **a**) and linear (part **b**) scales, respectively. Applied V_{DS} is 0.1 V. **c-f**, Statistical comparisons of transfer characteristics (part **c**) and their extracted effective mobility (part **d**), *SS* (part **e**), and V_T (part **f**) for a group of transistors measured before and after annealing. *p* values for one-way ANOVA are 0.23, 0.22, and 0.33, respectively, indicating no statistically significant difference.

Comment #2: *"The author demonstrated multi-tier devices using silicon nanomembrane junctionless transistors. Even though the transistors bypass the need for dopant activation that often requires high temperature, source/drain formation through epitaxy/implantation/annealing is still necessary to obtain low contact resistance. (In 2015, CEA Leti had demonstrated 3D VLSI with CoolCube process, in which solid phase epitaxy and nanosecond laser annealing are performed. VLSI IEEE, 2015. p. T48-T49.) Therefore, the reviewer is wondering whether the doping process can be performed before the transfer of silicon nanomembrane to define the source/drain of the transistors."*

Our response to comment #2: We thank the referee for raising this important and technically insightful point. We agree that low contact resistance is essential for high-performance devices and that, in principle, source/drain regions could be defined by heavy doping before transfer of the silicon nanomembrane. However, doing so would predefine the lateral footprint and locations of each transistor prior to transfer, making the vertical registration between tiers dependent on the mechanical alignment accuracy of the transfer process. As shown in Extended Data Fig. 7, this mechanical alignment is limited to the micron scale, which would severely constrain the achievable density of inter-tier vertical interconnects and the granularity of interlayer partitioning, both of which are central to monolithic 3D integration.



Extended Data Figure 7| Alignment of sacrificial access patterns between different tiers of transferred single-crystalline silicon nanomembranes. Optical images showing the misalignment less than 1 μm under low (part **a**, scale bar: 100 μm), medium (part **b**, scale bar: 50 μm), and high (part **c**, scale bar: 25 μm) magnifications. Inset: zoom-in view of the access patterns, scale bar: 5 μm. The white dashed circles as visual guide highlight the boundary of keep-out regions as part of the design rule in the PDK to avoid these holes.

In contrast, our junctionless device design employs a uniformly doped silicon nanomembrane, with source/drain regions defined lithographically after transfer without additional implantation, epitaxy, or annealing. This shifts what limits the alignment accuracy to lithography overlay, enabling sub-10 nm inter-layer registration (Fig. 2c). This high-precision alignment is essential for realizing dense, transistor-level vertical connectivity across tiers, which is one of the key advantages of monolithic 3D architectures.

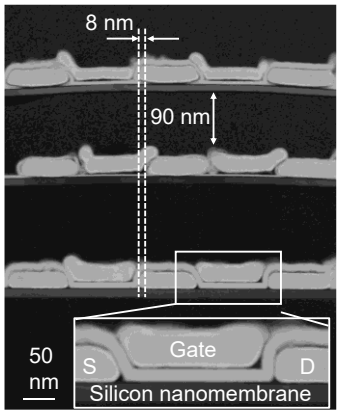


Figure 2. BEOL-compatible junctionless transistors built on transferred single-crystalline silicon nanomembranes. **c**, STEM micrograph showing three tiers of stacked junctionless transistor arrays separated by ~90 nm low-κ interlayer dielectrics. Scale bar: 50 nm. Inset: Zoom-in view of a single device.

We appreciate the reviewer's reference to the CoolCube process (CEA-Leti, VLSI 2015). While this work successfully demonstrated 3D VLSI integration, it relied on solid-phase epitaxy and nanosecond laser annealing at temperatures exceeding 600 °C. Such conditions are substantially above the 400 °C thermal budget required for BEOL-compatible processing. By avoiding such high-temperature steps, our approach enables fully BEOL-compatible monolithic 3D integration.

We also recognize the referee's concern regarding contact resistance in junctionless transistors built on a uniformly doped silicon nanomembrane. To address this, we have added an alternative

process flow to the revised manuscript showing how our approach can also yield conventional MOSFETs with raised, heavily doped source/drain to minimize contact resistance and an oppositely doped, recessed channel region, with added process complexity as tradeoff. These devices are fabricated from transferred silicon membrane with a vertically graded doping profile from a higher concentration near the surface to a lower concentration deeper inside the membrane. In this process, the source-drain and channel are still defined lithographically after transfer to ensure high inter-tier alignment accuracy, while avoiding post-transfer doping and activation to remain within a 400 °C thermal budget. Preliminary process and device results are included as new Extended Data Fig. 11 with details provided in Method in the revised manuscript (See below).

In summary, these additions clarify that while pre-transfer doping is feasible, it is fundamentally incompatible with the alignment precision required for dense monolithic 3D integration. Our junctionless architecture avoids this limitation, and our MOSFET demonstration shows that low-resistance source/drain engineering is also achievable within a $\leqslant$400 °C thermal budget.

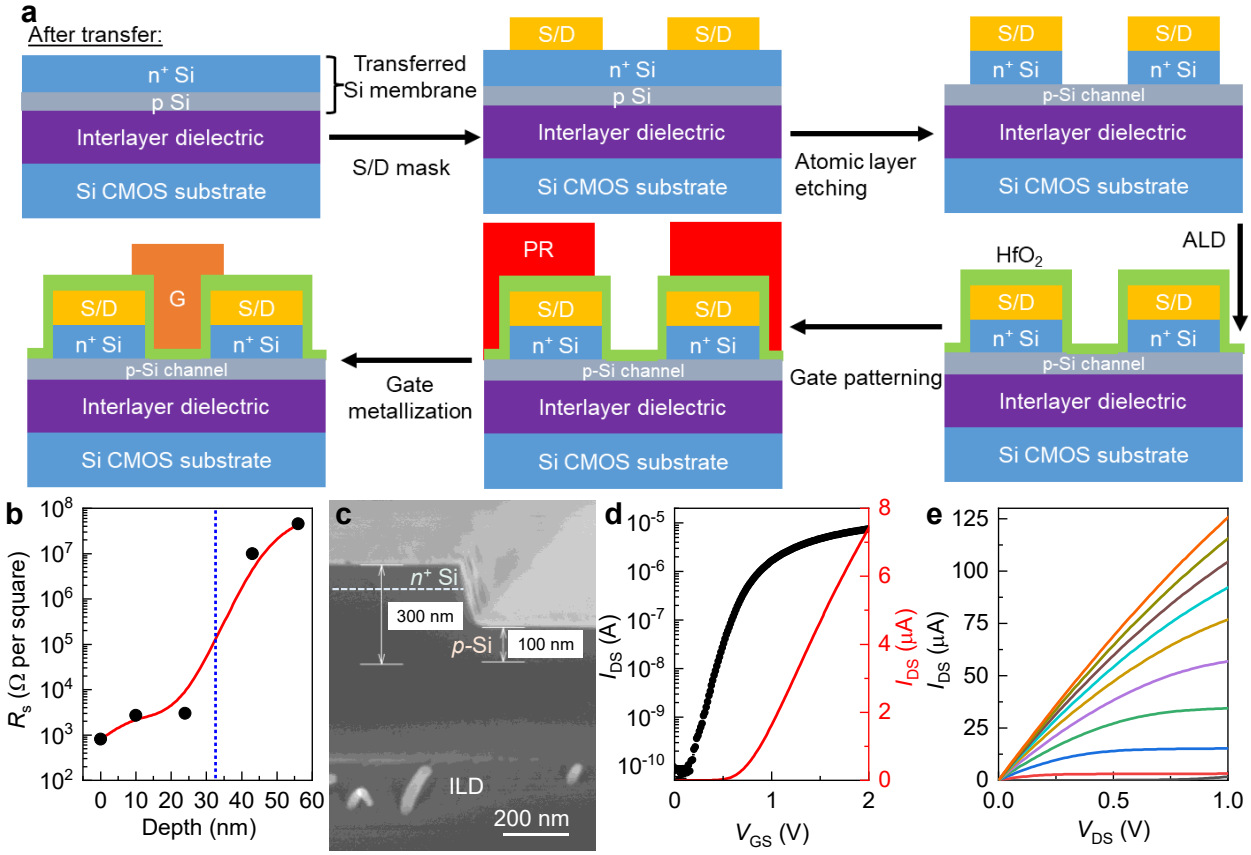
Our modification to the manuscript: We added the following statement in the revised manuscript: "*Because the source–drain regions are defined lithographically after each transfer, the registration accuracy is determined by lithography overlay rather than mechanical alignment during transfer, enabling inter-layer misalignment below 8 nm.*"

We also added: "*Meanwhile, MOSFETs incorporating raised, heavily doped source/drain regions and oppositely doped, recessed channels can also be fabricated from the transferred silicon nanomembranes within a <400 °C thermal budget (See Methods and Extended Data Fig. 11). Compared to junctionless transistors, such MOSFETs alleviate some limitations, including high contact and extension resistance as well as restricted threshold tunability, at the expense of a more complex fabrication flow.*"

Methods

Fabrication of silicon MOSFETs with raised source-drain and recessed channel: The process is schematically illustrated in Extended Data Fig. 11a. A SOI wafer with a 300 nm lightly *p*-doped (~10^{15} cm^{-3}) device layer was first blanket-coated with SOD (Filmtronics P505). Rapid thermal annealing was performed with a temperature ramp rate of 50 ^{o}C s^{-1} to 950 ^{o}C, followed by rapid cooling under flowing argon (1000 s.c.c.m.), producing a heavily *n*-doped surface layer ~35 nm deep (Extended Data Fig. 11b). After layer transfer, a AZ5214E photoresist film was spin-cast at 3000 rpm for 30 s. Alignment marks and source–drain patterns were defined in the resist, aligned to the mesh pattern of the transferred nanomembranes, and processed through development, sputtering of 0.4 nm Cr/60 nm Au, and lift-off in hot acetone. The resulting source–drain electrodes served as a hard mask for atomic-layer etching (ALE) of silicon using an Oxford mixed inductively coupled plasma (ICP)-RIE system. Each ALE cycle consisted of a purge step (RIE power was 0 W) and an etch step (RIE power was 30 W), with ICP power fixed at 750 W and gas flows of 40 s.c.c.m. SF_6 and 100 s.c.c.m. Ar. Each cycle removed ~1 nm of silicon with low surface roughness with the ALE smoothing effect[72]. After ~200 cycles, a 100 nm thick lightly *p*-doped recessed channel was formed, self-aligned to the raised, heavily *n*-doped source-drain and metal contacts (Extended Data Fig. 11c). A second

photolithography step patterned etch masks to protect the channel regions, followed by RIE to remove unprotected silicon, thereby defining device width and achieving device isolation. The resist was stripped in acetone, and an oxygen plasma clean was performed before depositing a 10 nm HfO_2 gate dielectric by ALD. A third photolithography step defined the gate electrodes, followed by sputtering of a 4 nm Cr/40 nm Au stack and lift-off in acetone. Finally, a fourth photolithography step opened contact windows for probing pads, and dry etching of the HfO_2 exposed the underlying contacts. Resist removal in acetone completed the transistor fabrication, all within a thermal budget $\leqslant 400\,°C$. The resulting devices exhibited effective mobilities exceeding 90 $cm^2V^{-1}s^{-1}$, as extracted from their transfer and output characteristics (Extended Data Fig. 11d and 11e).

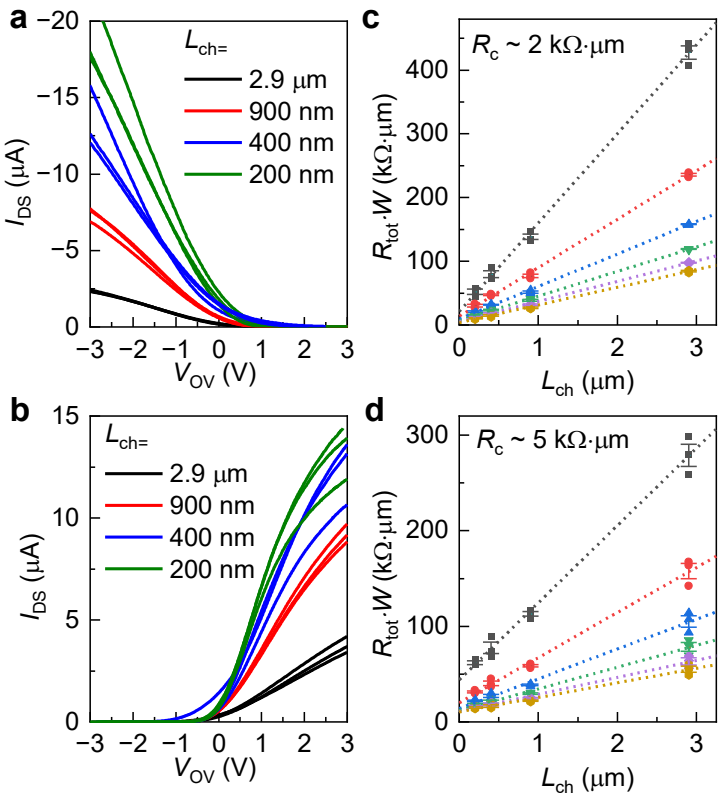


Extended Data Figure 11| MOSFETs fabricated on transferred silicon nanomembrane under a BEOL-compatible thermal budget. **a,** Schematic illustration of the fabrication flow. **b**, Evolution of silicon sheet resistance with ALE etching depth, measured by four-point probe. The blue dashed line serves as a visual guide to mark the approximate boundary position between the heavily doped surface layer and the oppositely and lightly doped channel. **c**, Cross-section SEM micrograph showing the heavily *n*-doped raised source-drain regions and light *p*-doped recessed channel defined by etching. **d**, Transfer characteristics of a completed MOSFET, plotted in both linear (red, right axis) and logarithmic (black, left axis) scales. Applied V_{DS}=0.1 V. **e**, Corresponding output characteristics. The device L_{ch}=3 μm and W=2 μm. The V_{GS} descends from 4.5 V at a step of 0.5 V from top to bottom.

Comment #3: *"TLM measurement tends to under/over-estimate the contact resistance during linear fitting of the data. The author should indicate the error bars of a set of data rather than fitting values extracted from a single device."*

Our response to comment #3: We thank the referee for raising this important point and agree that extracting contact resistance from a set of devices provides a more reliable assessment. In response, we repeated the TLM measurements on three devices for each channel length (12 devices in total) down to 200 nm and extracted the corresponding contact resistance values with statistical variation. The revised Extended Data Fig. 12 now reports these results with error bars representing the standard deviation across the device set, thereby capturing device-to-device variability. The overall conclusion remains unchanged: the contact resistance is consistently higher for *n*-FETs than for *p*-FETs. We have incorporated the updated data into the revised manuscript.

Our modification to the manuscript: We incorporated error bars and data points extracted from a set of devices:



Extended Data Figure 12| Contact resistance of low-temperature fabricated silicon junctionless transistors. a-b, Transfer characteristics of *p*-FETs (part **a**) with platinum contact and *n*-FETs (part **b**) with Sb/Ti contact, with L_{ch} of 2.9 μm (black), 900 nm (red), 400 nm (blue), and 200 nm (green), respectively. **c-d**, Width-normalized device total resistance (R_{tot}·W) as a function of L_{ch} at varied gate overdrive (V_{OV}=V_G–V_T) from ±1 V to ±2.5 V from top to bottom for *p*-FETs (part **c**) and *n*-FETs (pard **d**), respectively. The y-axis intercepts give the parasitic contact resistance $2R_C$ as extracted based on the transmission-line method. Dashed lines represent linear fittings to the data. Error bars represent s.d..

Comment #4: *"Vdd of 3V is considered too high for a silicon CMOS circuit. Suitable Vdd for a silicon device at legacy node typically ranging from 0.9V – 2V. The reviewer suggests the author to show the voltage-transfer characteristics of the devices in suitable Vdd."*

Our response to comment #4: We thank the referee for this helpful suggestion and agree that demonstrating circuit operation at suitable V_{DD} is important. Following the recommendation, we have measured and replotted the voltage-transfer characteristics of CMOS inverters, NOR and NAND gates, and SRAM cells at V_{DD} values of 0.7, 1.0, 1.5, and 2.0 V, as shown in the revised Fig. 4. The circuits exhibit correct logic functionality and well-defined voltage-transfer behavior across this typical supply-voltage range. Specifically, the inverter shows a high gain of ~ 40 V/V and full swing outputs at V_{DD} of 0.7 V in the voltage-transfer characteristics.

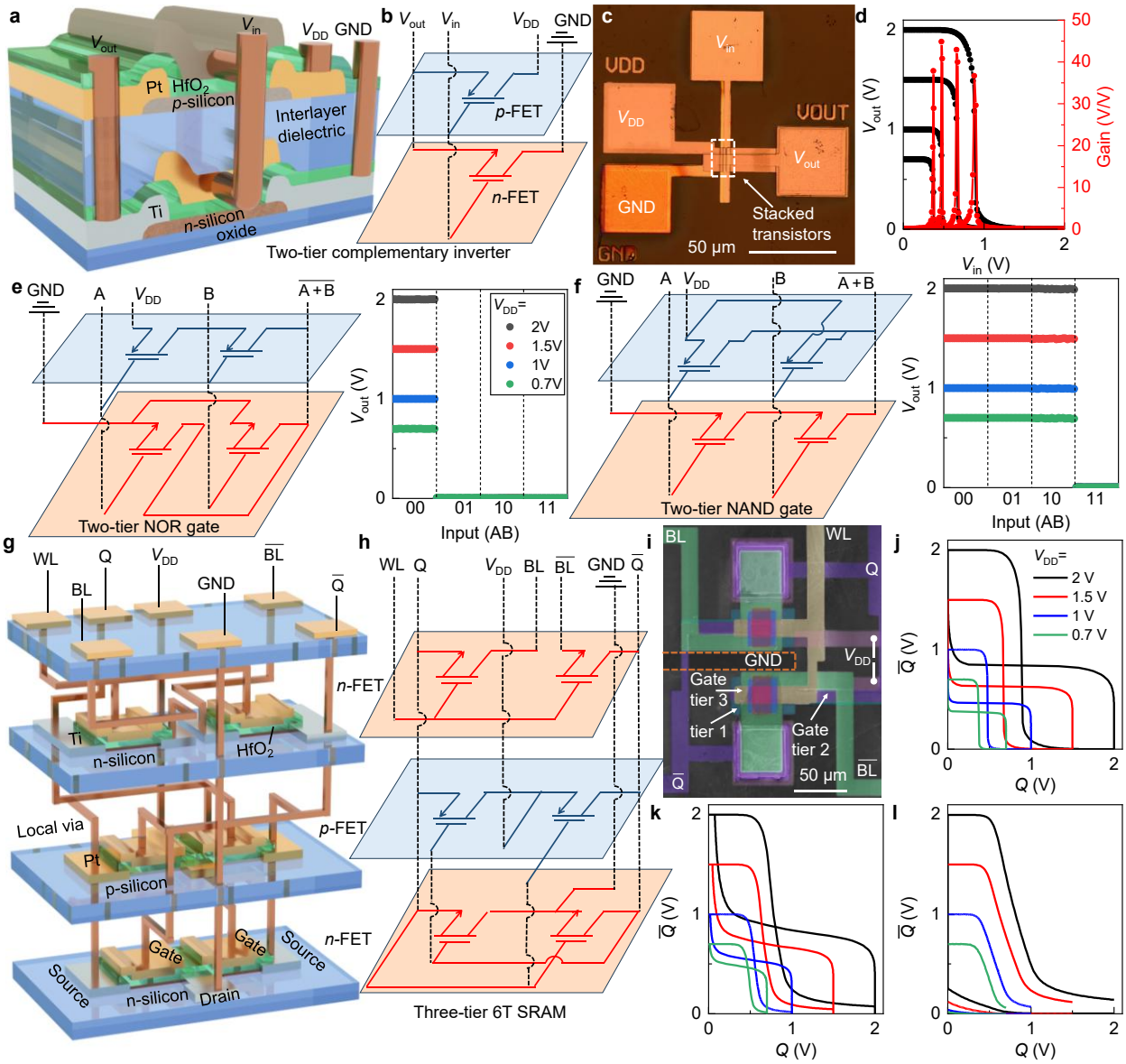


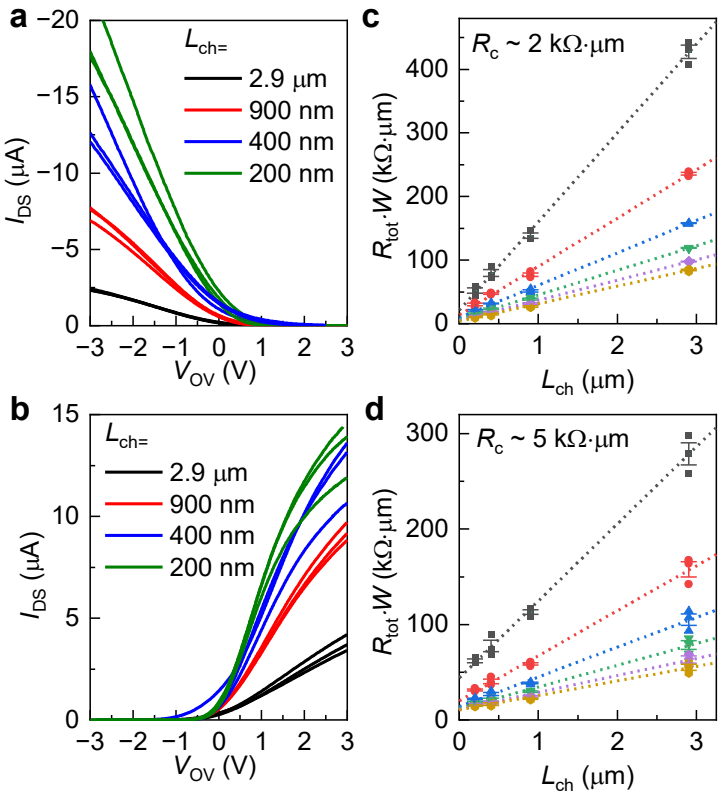
Figure 4. Monolithic 3D integrated logic circuits built on BEOL-compatible silicon junctionless transistors. a-d, Layout schematic (part **a**), circuit diagram (part **b**), optical micrograph (part **c**), and voltage-transfer characteristics (part **d**, left axis) and the associated voltage gain (right axis, red) of the monolithic 3D integrated inverter extracted with different V_{DD}. **e-f**, Circuit diagrams and output-input characteristics of a 3D NOR gate (part **e**) and a 3D NAND gate (part **f**). Applied V_{DD} ranges from 2 V (black) to 1.5 V (red), 1 V (blue), and 0.7 V (green). **g-i**, Layout schematic (part **g**), circuit diagram (part **h**), and false-coloured SEM micrograph (part **i**) of a monolithic 3D SRAM cell. **j-l**, Hold (part **j**), read (part **k**), and write (part **l**) margin measurements with V_{DD} of 2 V (black) to 1.5 V (red), 1 V (blue), and 0.7 V (green), respectively. Scale bars: 50 μm.

Our modification to the manuscript: We have updated Fig. 4 in the revised manuscript with appropriate V_{DD} ranges and modified associated discussions: "*We first constructed an inverter by connecting a p-FET sitting directly above a n-FET, with its layout schematic, diagram, and micrograph illustrated in Figs. 4a-c. Its voltage-transfer characteristics show a gain of 36–45 V/V with a full swing to V_{DD} between 0.7–2 V (Fig. 4d).*"

Comment #5: *"The reported p-type junctionless Si FETs deliver higher current density than their n-type counterparts. While this might seem counterintuitive based on the typical electron vs hole mobility in silicon. The authors should provide a more detailed discussion of these mechanisms to clarify why the p-FETs outperform the n-FETs in this work."*

Our response to comment #5: We thank the referee for raising this important point. Although p-type junctionless Si FETs outperforming n-type devices may appear counterintuitive based on bulk electron–hole mobility, this behavior is in fact expected in highly doped, ultrathin junctionless channels, where contact resistance and mobility-degradation asymmetries dominate. Three mechanisms act synergistically:

- **Lower parasitic contact resistance in *p*-FETs:** As shown in Extended Data Fig. 12, the *p*-FETs with Pt contacts exhibit more than twofold lower parasitic contact resistance than the *n*-FETs with Ti contacts (2 kΩ·μm *versus* 5 kΩ·μm). This higher series resistance in the *n*-FETs directly limits their drive current and can outweigh the intrinsic mobility advantage of electrons.



Extended Data Figure 12| Contact resistance of low-temperature fabricated silicon junctionless transistors. a-b, Transfer characteristics of *p*-FETs (part **a**) with platinum contact and *n*-FETs (part **b**) with Sb/Ti contact, with L_{ch} of 2.9 μm (black), 900 nm (red), 400 nm (blue), and 200 nm (green), respectively. **c-d**, Width-normalized device total resistance (R_{tot}·W) as a function of L_{ch} at varied gate overdrive (V_{OV}=V_G–V_T) from ±1 V to ±2.5 V from top to bottom for *p*-FETs (part **c**) and *n*-FETs (pard **d**), respectively. The y-axis intercepts give the parasitic contact resistance $2R_C$ as extracted based on the transmission-line method. Dashed lines represent linear fittings to the data. Error bars represent s.d..

- **Reduced electron-hole mobility disparity at high doping**: While electrons in lightly doped silicon (10^{15} cm^{-3}) have ~4.7× higher mobility than holes (1331 $cm^2V^{-1}s^{-1}$ *versus* 283.5 $cm^2V^{-1}s^{-1}$at 300 K), this disparity narrows substantially at the high doping levels required for junctionless transistors, with the carrier transport limited by ionic scattering. At the average doping concentration employed in our devices ($2×10^{19}$ cm^{-3}), the electron mobility decreases to ~97 $cm^2V^{-1}s^{-1}$, while the hole mobility remains ~57 $cm^2V^{-1}s^{-1}$, a difference of only ~1.7×. Under these conditions, both carriers are strongly limited by ionized impurity scattering, reducing the intrinsic mobility advantage of electrons.

- **Stronger electron mobility degradation in ultra-thin silicon channels:** The above values are for bulk silicon. In ultra-thin silicon bodies (<10 nm), electron mobility is more strongly degraded by surface roughness scattering than hole mobility. For example (see Fig 5 and Fig. 6 in 10.1109/IEDM.2000.904408), reducing silicon thickness from 54 nm to 5.2 nm decreases electron mobility by ~34% (from ~650 to ~430 $cm^2V^{-1}s^{-1}$), whereas hole mobility decreases by only ~18% (from ~170 to ~140 $cm^2V^{-1}s^{-1}$). This differential degradation further narrows the electron-hole mobility gap in our ultra-thin junctionless channels.

Taken together, these three effects—lower contact resistance in *p*-FETs, reduced electron-hole mobility disparity at high doping, and stronger electron mobility degradation in ultra-thin channels—fully reconcile the apparent contradiction and explain why our *p*-type junctionless transistors achieve comparable, and in this case slightly superior, current density relative to their *n*-type counterparts. Similar behavior has also been observed in previous works on junctionless transistors (Fig. 4b in 10.1109/VLSIT.2018.8510705, I_{on} ~260 μA/μm for *p*-FET and ~230 μA/μm for *n*-FET). We have incorporated this detailed discussion into the revised manuscript.

Our modification to the manuscript: We added the following statement in the revised manuscript: "*In junctionless transistors, the high channel doping concentration (>10^{19} cm^3) narrows the disparity between electron and hole mobilities, as both become strongly limited by ionic impurity scattering*[60]. *In addition, ultrathin silicon body (<10 nm) leads to more pronounced mobility degradation for electrons, which are more susceptible to surface roughness scattering than holes*[61]. *The slightly higher performance of p-FETs, which is also reported in previous studies of junctionless devices*[62]*, is further attributed to their lower parasitic contact resistance relative to n-FETs (Extended Data Fig. 12), an effect that outweighs the intrinsic mobility advantage of electrons.*"

Referee #1:

Summary comments: *"I have carefully reviewed the revised version of the manuscript and the authors' response to the previous review comments. The authors have made substantial efforts to address the major concerns raised in the initial review. The manuscript has improved notably in clarity, methodology, and presentation."*

Our response: We sincerely thank the referee for the positive assessment and for the constructive feedback that helped us improve the clarity and presentation of our manuscript.

Comment #1: *"Summary of the key results: After the major revisions, the abstract is now clearly formulated and provides a concise and understandable overview of the main findings. The key results are presented in a coherent manner, improving the accessibility and impact of the manuscript. I appreciate that the authors have followed my suggestion to reduce the number of references and make them more specific. While the abstract has improved significantly, one more sentence about the roll-to-roll transfer-printing process and it's practical relevance would make it even more complete."*

Our response to comment #1: We thank the referee for the thoughtful suggestion. Following this recommendation, we added a sentence to the abstract that highlights the roll-to-roll transfer-printing process and its practical relevance to scalable monolithic 3D integration.

Our modification to the manuscript: We included the following sentence in the revised abstract of the manuscript: "*Here we show that uniformly doped, ultrathin (≤10 nm) single-crystalline silicon nanomembranes can be vertically stacked using a roll-transfer-printing process that is scalable to wafer-scale and tolerant to substrate topology and surface roughness, enabling multi-tiers of complementary junctionless transistors to be sequentially fabricated on the same starting substrate under a processing temperature ≤ 400 ^{o}C.*"

Comment #2: *"As stated in my initial review, I consider the work to be both original and significant, addressing an important topic in the field."*

Our response to comment #2: We thank the referee for this positive comment and assessing our work as "*both original and significant, addressing an important topic in the field*".

Comment #3: *"The methodology appears sound and appropriate for the research question. The data quality is adequate, and the results are presented clearly with well-prepared diagrams and figures."*

Our response to comment #3: We thank the referee for this encouraging feedback and appreciate the recognition of our methodological approach and improvements to the clarity of our figures, with referee's help.

Comment #4: *"Appropriate use of statistics and treatment of uncertainties: The authors show a strong awareness of the statistical relevance of their measurements. In most figures, results are presented for ensembles of devices, often using boxplots or histogram-style representations. This provides a clear picture of device-to-device variability and indicates a careful treatment of*

uncertainties."

Our response to comment #4: We thank the referee for this positive comment regarding our statistical analysis and presentation of device variability.

Comment #5: *"Conclusions: robustness, validity, reliability: The authors have added a sentence addressing the relatively modest performance of the devices, which helps clarify this aspect of the results. The authors have improved the conclusion section in the revised version. The conclusions are now clearly stated and better supported by the presented results. I am satisfied with the authors' final interpretation, and this aspect of the manuscript can be considered adequately addressed."*

Our response to comment #5: We thank the referee for the constructive feedback that helped strengthen our conclusion.

Comment #6: *"Suggested improvements: experiments, data for possible revision. The major points raised in the initial review have been sufficiently addressed. I have no further requests for additional experiments or data, and no further revision is required from my side. The authors briefly mention a possible upscaling of their approach in the response letter. I find this aspect very interesting and promising, as it could strongly increase the practical relevance of the work. Although it may not be essential for the current manuscript, it represents an exciting direction for future research."*

Our response to comment #6: We thank the referee for assessing our revisions as satisfactory. We agree that scaling up our approach is an exciting direction, and we are actively pursuing this path in collaboration with our industry partners.

Comment #7: *"References: appropriate credit to previous work? The references are appropriate and reflect the state of the art quite well. One minor revision point is to explicitly mention the applied transfer technique either in the abstract and in the list of keywords. In my understanding, this technique represents the key methodological approach of the work and should therefore be highlighted more prominently."*

Our response to comment #7: We thank the referee for noting that our references appropriately reflect the state of the art. Following the suggestion, we have explicitly highlighted the roll-transfer-printing technique in the abstract of the revised manuscript.

Our modification to the manuscript: We included the following sentence in the revised abstract of the manuscript: "*Here we show that uniformly doped, ultrathin (≤10 nm) single-crystalline silicon nanomembranes can be vertically stacked using a roll-transfer-printing process that is scalable to wafer-scale and tolerant to substrate topology and surface roughness, enabling multi-tiers of complementary junctionless transistors to be sequentially fabricated on the same starting substrate under a processing temperature ≤ 400 ºC.*"

Referee #2:

Summary comments: *"Thank you very much for the detailed reply to the questions raised in my previous evaluation. All my concerns were addressed in a clear, honest, and professional manner. I particularly appreciate the authors' inclusion of additional variability data and their transparent discussion of the underlying sources of variability linked to integration limits, as well as the correlation with semiconductor and oxide thickness.*

Similarly, the response regarding channel-length scaling was thorough. The added TCAD simulations—showing a scaling limit near ~30 nm and the need for more advanced architectures beyond that point—provide valuable context and further strengthen the credibility of the work.

Overall, I am satisfied with the clarifications and the additional data provided. I recommend the manuscript for publication as it is."

Our response: We sincerely thank the referee for the positive evaluation and for recommending our manuscript for publication in *Nature* "*as it is*". We greatly appreciate the referee's thoughtful feedback throughout the review process, which helped us further strengthen the clarity and rigor of the manuscript.

Referee #3:

Summary comments: *"The authors have addressed some of the technical issues and provided explanation or justification through experimental data. However, some of the following points remain and need to be addressed."*

Our response: We thank the referee for the thoughtful comments that helped us improve the quality and clarity of our manuscript. Detailed, point-by-point responses to each of the remaining concerns are provided below.

Comment #1: *"Area loss due to etch holes: The authors have added explanation and supporting data to show Si layer release can be achieved with low density (500 µm pitch) holes. This is satisfactory for the current status and provides readers a context."*

Our response to comment #1: We thank the referee for helping us improve the clarity of our presentation and for assessing the new data on low release-hole density as satisfactory.

Comment #2: *"Layer transfer alignment: This point was raised previously and is a drawback of the method. The authors have also included a statement "These dead space patterns can be aligned across tiers during transfer using a mechanical stage, with micron-scale resolution". Hence, direct wafer bonding remains free of any overlay errors and better than the proposed method for high performance logic, especially when 3D circuits are needed."*

Our response to comment #2: We thank the referee for acknowledging our revision and have more explicitly highlighted this limitation of our approach in the revised manuscript.

Our modification to the manuscript: We added the following statement to the revised manuscript: "*These dead-space patterns can be aligned across tiers during transfer using a mechanical stage; however, the micron-scale alignment resolution limits the extent to which area loss can be reduced solely by decreasing the access-hole diameter (Fig. S4).*"

Comment #3: *"Comparison with direct wafer bonding: Authors claim multi-layer stacking beyond two layers as unique advantage. Direct wafer bonding (DWB) layer transfer has fundamentally no restriction as well for multi-layer transfer. More than 2 layers have been demonstrate in hybrid bonding (which uses same mechanism as DWB with added complexity of embedded interconnects (https://doi.org/10.1109/TCPMT.2025.3533926; DOI 10.1109/ECTC51529.2024.00106). As example of only layer transfer (without embedded interconnects) 3 tier has also been demonstrated (https://doi.org/10.1109/SOI.2007.4357865). This aspect is not novelty of the proposed method. DWB and hybrid bonding demonstrations do not use thin dielectrics between layers as most of those involve advanced CMOS where inter-tier metals are needed. It is not the limitation of the method but a technological choice. Same would apply to proposed method. Any advanced CMOS demonstration would require inter-tier metallization, technologically relevant gate height and other modules that would necessitate a thicker inter-tier dielectric. Regarding the cost argument, the claims can not be fully substantiated and also compared between an academic instrument/process and 300 mm industrial grade process. Only feature that remains unique to the process compared to DWB is*

the ability to transfer over topography, which in the revision hasn't be expanded upon. The authors should remove the sentences that claim the process allows more tiers compared to low-temperature DWB."

Our response to comment #3: We thank the referee for the detailed clarification and agree that direct wafer bonding (DWB) is not "*fundamentally*" restricted from multilayer stacking. However, its practical implementation becomes challenging due to limited tolerance to substrate topography and surface roughness during vertical layer scaling, an aspect that represents a key advantage of our transfer-printing approach as the referee agrees with.

The multilayer demonstrations cited by the referee involve stacking prefabricated dies or wafers, rather than monolithically integrating multiple tiers of single-crystalline silicon transistors on the same starting substrate. To our knowledge, monolithic stacking of more than two tiers of single-crystalline silicon transistors has not been previously demonstrated, and our process therefore represents a significant technological advance in this context. Thin interlayer dielectrics make it easier to form inter-tier interconnects with low aspect ratio, as what we used in our monolithic 3D circuit demonstrations.

We also appreciate the referee's point regarding cost analysis and agree that a full, quantitative comparison is beyond the scope of the current manuscript. In accordance with the referee's recommendation, we have revised the text to avoid discussing the number of monolithic transistor layers achievable with low-temperature DWB and have focused solely on factual statements regarding the demonstrated capabilities of our process.

Our modification to the manuscript: We removed comparison with low-temperature direct wafer bonding in the revised manuscript, with only factual statement that: "*Owing to its process simplicity, transfer can be readily repeated to realize a multiple (i.e., >2) tier stack.*"

In the revised manuscript, we added the following statement to acknowledge that: "*The full cost analysis and comparison in a manufacturing setting is out of the scope of the current manuscript.*"

Comment #4: *"Authors have demonstrated layer transfer for bulk wafer with their method. Unfortunately, it adds lithography layers which is adds cost and is leads additional overlay requirement. While the data is necessary to show utilization of the method with bulk wafers, it is not equivalent to reusability as in SOI wafer which is full layer transfer without the need for lithography on donor wafer. Moreover, Si (111) wafers are not standard. Please add AFM scan showing the thickness variation across the layer transferred from bulk wafer (show the variation closer to the edge of the pattern)."*

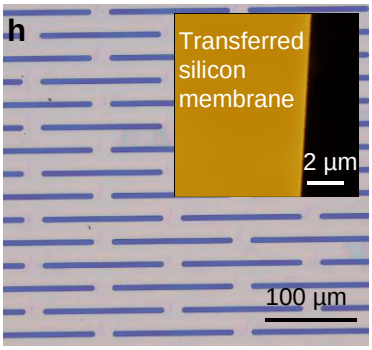
Our response to comment #4: We thank the referee for acknowledging the value of the added data demonstrating the applicability of our method to bulk silicon wafers. We agree that a full cost and yield comparison with DWB-based processes is complex: our approach requires lithography on the donor wafer, which could be low cost and high yield with only micron resolution required, whereas SmartCut-based DWB involves ion implantation, wafer bonding, and polishing with limited tolerance to substrate topography and surface roughness. We now explicitly acknowledge this in the revised manuscript.

We also agree that (111) wafers are less commonly used in logic device fabrication. However, they are readily available at low cost and are used in fabricating bipolar junction transistors. (111) wafers also offer potential advantages such as enhanced hole mobility and therefore are a viable substrate choice for our demonstration and future monolithic 3D silicon technology.

As suggested by the referee, we have added an AFM scan showing the uniform thickness of silicon nanomembranes derived from bulk wafers, starting from the edge of the patterned region.

Our modification to the manuscript: In the revised manuscript, we added the following statement to acknowledge: "*The full cost analysis and comparison in a manufacturing setting is out of the scope of the current manuscript.*"

We have added an AFM scan showing the uniform thickness of silicon nanomembranes derived from bulk wafers, starting from the edge of the patterned region:



Extended Data Figure 2| Preparation of freestanding silicon nanomembranes from bulk silicon (111) wafers. g-i, Optical photo (part **g**), top-view optical micrograph (part **h**, inset: AFM micrograph showing a uniform film from the edge of the release-trench pattern), and infrared image (part **i**) of an nanomembrane derived from a bulk (111) silicon wafer and transferred to a SiO_2/Si subtracted covered by spin-cast low-κ ILD.

Comment #5: *"The method is ok for research approaches."*

Our response to comment #5: We thank the referee for assessing our method as suitable for research-level applications.

Comment #6: *"Donor wafer CMP is only needed for recycling and specifically the need for high volume manufacturing to reduce costs. The proposed method from bulk wafers requires lithography and adds overlay overhead and is not applicable for full layer transfer (unlike SmartCut). Hence the claimed benefits are not justified. Oxidation based thinning of Si layer is fairly standard in SOI CMOS flow and is also used for SmartCut like process. Non-uniformities, contamination risks also exist in oxidation based process as well. Since the top-Si thickness variations and contamination levels across SmartCut SOI wafers is the same as standard bulk wafers (both being used in high-volume CMOS manufacturing), the risks claimed due to CMP are clearly not show stoppers. The authors should modify the sentences to put their work context for research and low volume prototype."*

Our response to comment #6: We thank the referee for acknowledging the respective benefits and limitations of our process relative to SmartCut-based DWB. Our approach requires

lithography on the donor wafer, which could be low cost and high yield with only micron resolution required, and introduces a minimum area-loss overhead, whereas SmartCut-based DWB relies on ion implantation, wafer bonding, and chemical-mechanical polishing (CMP) to define the transferred layer thickness, and has limited tolerance to substrate topography and surface roughness. Our method is not intended to replace SmartCut-based DWB or to offer universal advantages; rather, it provides an alternative pathway that is particularly well-suited for research-level and low-volume prototyping, consistent with the referee's recommendation. We have emphasized this context in the revised manuscript.

Regarding CMP, our intention was not to imply that CMP poses a showstopper for SmartCut-based DWB, but to note that avoiding CMP on the transferred film removes potential sources of particle contamination and within-wafer non-uniformity. Oxidation, as a front-end-of-line process, generally exhibits higher uniformity and lower contamination risk than back-end-of-line polishing. We have clarified this point in the revised text.

Our modification to the manuscript: We added the following statement in the revised manuscript: "*This wafer-scale transfer-printing process offers several advantages that are especially valuable for research and low-volume prototyping compared with stacking silicon films based on wafer bonding followed by wafer thinning".*

We added "*potential*" in the statement of "*This avoids potential risks of particle contamination and within-wafer non-uniformity associated with pad wear and slurry flow variations in chemical-mechanical polishing".*

Comment #7: *"Defectivity characterization is sufficient"*

Our response to comment #7: We thank the referee for the constructive feedback that helped us strengthen the defectivity characterization and for assessing the resulting analysis as sufficient.

Comment #8: *"The authors have attempted to demonstrate junction-less flow as competitive. The issue is access and contact resistance and these can not be mitigated in FinFET geometry. In fact, they become worse. It is more challenging to obtain higher drive current in FinFET compared to planar geometry. The authors demonstrate layer transfer with doped source/drain layers. The same method can also be used in DWB method of 3D monolithic and is not unique benefit of proposed method. The approach leads to significantly overlapped transistor design (high gate-source/drain capacitance) and channel damage during doped layer etch. Hence it is not a useful process option. Post-layer transfer, doped source/drain in a self-aligned process is a necessity and is not solved by the process shown by authors. These details do not address the issue and justify the claim of high performance with low temperature.*

Also, the sentences added by authors seem to indicate planar MOSFETs have limited drive current than FinFETs. This is an incorrect assumption without all parameters factored in (short-channel effects, footprint, necessary overdrive and supply voltage etc.). The authors should remove the sentences "as limited mainly by the planar channel structure. Future iterations that transition from planar channels to vertical fins are expected to significantly narrow this performance gap".

Overdrive conditions: Constant overdrive comparison is used to compare at equivalent field for various EOT. The authors have misunderstood the concern. It is expected that EOT of

academic devices is lower than that of state-of-the art CMOS. However, constant overdrive comparison allows current and charge density in the channel comparison across these. The benchmark table should show ION at fixed overdrive conditions across technologies. The figure 14b is misleading since it compares ION and various conditions across technologies and should be removed.

Benchmarking: The compared Si DWB benchmark references utilizes 525°C because of the raised S/D module (RSD). As mentioned in the previous point, state-of-the art transistor architecture requires non-overlapping flow. The bonding and transfer process itself requires temperature below 400°C. This context should be mentioned."

Our response to comment #8: We appreciate the referee's detailed assessment. We agree that access and contact resistance is a challenge for junctionless transistors and cannot be mitigated by transitioning to a FinFET geometry. Our intention was not to imply otherwise. While FinFETs can, in principle, increase drive-current density through a larger effective channel width within the same device footprint with their vertical fins, we agree that the practical benefits depend on multiple engineering factors, as the referee notes. To avoid any confusion, we have removed the speculative statement regarding future fin-based iterations from the revised manuscript, following referee's recommendation.

We note that a self-aligned gate process with raised source/drain regions can be implemented at $\leqslant 400\,^{\circ}\mathrm{C}$ as a natural extension of our approach, and we include a schematic of such a low-temperature process flow (patent pending) to illustrate a viable path forward. We plan to experimentally verify this approach in future work. The channel damage is minimum after mild atomic layer etching, as evident from the high field-effect mobility> 90 $cm^2V^{-1}s^{-1}$ we extracted from our fabricated devices in experiment.

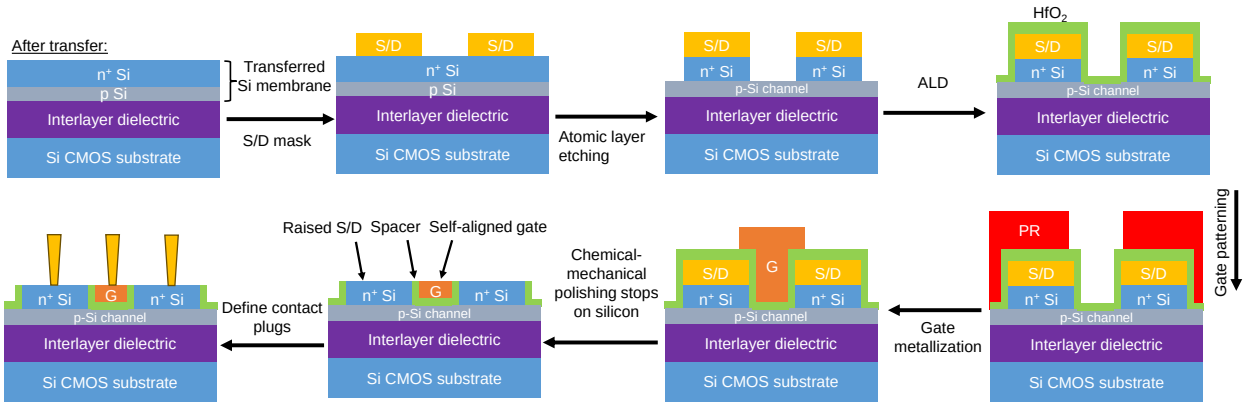


Fig. 1 Schematic of a low-temperature process to fabricate MOSFETs with a self-aligned gate and raised source/drain on a transferred silicon membrane.

For benchmarking, constant overdrive will not give the same gate electric field for devices with different EOTs or gate oxide physical thicknesses. The applied gate overdrive and EOT jointly determine the inversion charge density $Q_{inv} \approx C_{ox}(V_g - V_T) = \frac{\varepsilon_{SiO_2}}{EOT}(V_g - V_T)$ and therefore the achievable I_{on}. Thus, comparing I_{on} in a way that reflects carrier transport requires adjusting the applied overdrive for devices with different EOTs. In addition, because many prior publications do not report I_{on} at all overdrive values, a strict constant-overdrive comparison

across technologies is often not feasible. Our approach, which compares the maximum I_{on} reported under the overdrive appropriate for each device's EOT, provides qualitative insight while acknowledging these limitations. To avoid confusion, we have explicitly stated the comparison condition in the revised caption.

We thank the referee for pointing out the context of the Si DWB reference. We now explicitly note that the reported 525 °C thermal budget arises from the epitaxial raised source/drain module, while the bonding step remains below 400 °C.

Our modification to the manuscript: We revised the sentence to: "*Their output current density remains below that of FEOL MOSFETs in more advanced technology nodes, which surpass 1 mA·μm^{-1} with smaller device dimensions and 3D channels*", removing "*as limited mainly by the planar channel structure. Future iterations that transition from planar channels to vertical fins are expected to significantly narrow this performance gap"* as suggested by the referee.

In the revised manuscript, we added the following statement to emphasize minimum damage to silicon channel by ALE: "*The resulting devices exhibited effective mobilities exceeding 90 cm^{2}V^{-1}s^{-1}, as extracted from their transfer and output characteristics (Extended Data Fig. 4d and 4e), validating minimum damage to remaining silicon by ALE.*"

In the revised manuscript, we added the following sentence to acknowledge the possibility of forming self-aligned gate with the help of CMP: "*A chemical-mechanical polishing stopping on the raised silicon source-drain can form self-aligned gate with the HfO_2 dielectric also serving as spacers.*"

In the caption for Extended Data Fig. 5, we added: "*The maximum I_{on} was obtained under the maximum overdrive applied which differs across technologies due to different EOTs.*"

In the revised manuscript, I added: "*.... monolithic 3D silicon transistors formed by wafer bonding with a higher thermal budget of 525 ^oC for forming epitaxial raised source-drain contacts*".

Comment #9: *"DIBL is not close to simulation value. In CMOS technology terms it is a large offset 11 mV/V vs 60 mV/V. Modify the sentences. Subthreshold slope in high Vd is not mentioned."*

Our response to comment #9: We thank the referee for this helpful suggestion and have revised the discussion accordingly. Our measurements show no degradation of the subthreshold slope at high V_{DS}, and we have added this description to the revised manuscript. We also clarified the comparison between measured and simulated DIBL values as recommended.

Our modification to the manuscript: In the revised manuscript, we stated that: "*These short-channel junctionless transistors exhibited a small drain induced barrier lowering (DIBL) of 60±10 mV/V, with no degradation of the SS under high V_{DS}. The deviation from TCAD predictions of 11.6±0.5 mV/V is attributed to silicon-HfO2 interface traps (Fig. S9).*"

Comment #10: *"Cost and manufacturability: Authors have not done any cost model comparison to claim the benefits. Particularly, the proposed layer transfer from bulk wafers adds lithography steps (that are known to be expensive and causing yield issues) and adds overlay errors.*

Therefore, it is not equivalent to wafer-bonding layer transfer. Complementarity: While the proposed process allows layer-transfer on topography, and relevant for research (as mentioned by authors), it should reflect in the abstract and main claim of the paper."

Our response to comment #10: We thank the referee for raising this important point. We agree that a full cost-model analysis is beyond the scope of the present manuscript, and we have clarified this explicitly in the revised text.

In accordance with the referee's recommendation, we have also revised the abstract to highlight the unique capability of our process to perform layer transfer over substrate topography and to emphasize its relevance for research-level and low-volume prototyping applications.

Our modification to the manuscript: In the revised manuscript, we added the following statement to acknowledge: "*The full cost analysis and comparison in a manufacturing setting is out of the scope of the current manuscript.*"

In the revised abstract, we added: "*Here we show that uniformly doped, ultrathin (≤10 nm) single-crystalline silicon nanomembranes can be vertically stacked using a roll-transfer-printing process that is scalable to wafer-scale and tolerant to substrate topology and surface roughness, enabling multi-tiers of complementary junctionless transistors to be sequentially fabricated on the same starting substrate under a processing temperature ≤ 400 °C*" and "*Our demonstrations provide a promising route toward silicon-based monolithic 3D circuits, especially for research and low-volume prototyping.*"

Comment #11: *"The proposed work's novelty is in transfer process. The subsequent MOSFET process flow does not feature novel process or conditions that could make them high performance. Low temperature process (400°C) is due to the choice of the process flow that does not include process modules necessary for higher performance CMOS. The challenges for <=400°C are due to various CMOS process modules (spacers, RSD, high-k/metal gate etc.) which are integral part of advanced CMOS flow (authors can refer to extensive studies done on Si 3D Monolithic integration with DWB).*
Authors could have improved the manuscript by adding relevant device data (based on high topography features) to demonstrate the clear benefits of the proposed layer transfer."

Our response to comment #11: We thank the referee for acknowledging that the primary novelty of this work lies in the transfer-printing process, which is indeed the central contribution of the manuscript.

Our subsequent demonstrations of junctionless and MOSFET devices, which incorporate advanced features such as raised source/drain (RSD) regions and high-k/metal-gate stacks, required non-trivial device-structure and process-flow adaptations to achieve record-high performance under the stringent $\leqslant 400\,^{\circ}\mathrm{C}$ back-end-of-line compatible thermal constraint, as a significant technology milestone. These demonstrations are intended to validate the electrical integrity of the transferred nanomembranes and to illustrate the vertical scalability of multi-tier monolithic 3D silicon transistor integration within this thermal budget.

We appreciate the referee's suggestion to further explore devices fabricated on silicon nanomembranes transferred onto high-topography features. This is an important direction, and we are actively pursuing such studies in ongoing work, including investigations of strain effects and device behavior on engineered topography, which will be reported in separate manuscripts.

Referee #4:

Summary comments: *"Overall, the responses are clear and technically sound, and the revisions strengthen the manuscript without introducing new ambiguities. I therefore recommend acceptance of the manuscript in its current form."*

Our response: We sincerely thank the referee for the positive evaluation and for recommending our manuscript for publication in *Nature* "*in its current form*". We greatly appreciate the referee's thoughtful feedback throughout the review process, which helped us further strengthen the clarity and rigor of the manuscript.