

2026 37th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC) | 979-8-3315-8050-6/26/\$31.00 ©2026 IEEE | DOI: 10.1109/ASMC69324.2026.11551272

Authorized licensed use limited to: Nova Ltd.. Downloaded on June 16, 2026 at 21:37:48 UTC from IEEE Xplore. Restrictions apply.

and repeated N sweeps, the live time is $T \equiv N\tau$ and the observed counts satisfy

$$c_k \sim \text{Poisson}(r_k T), \quad \hat{r}_k \equiv \frac{c_k}{T} \quad (\text{counts s}^{-1}).$$

Because Poisson fluctuations obey $\text{Var}(c_k) = \mathbb{E}[c_k] = r_k T$, the rate (counts/s) estimator is unbiased with variance that falls inversely with live time: [1], [2], [7]

$$\mathbb{E}[\hat{r}_k] = r_k, \quad \text{Var}(\hat{r}_k) = \frac{r_k}{T}, \quad (1)$$

$$\text{RSD}(\hat{r}_k) = \frac{\sqrt{\text{Var}(\hat{r}_k)}}{\mathbb{E}[\hat{r}_k]} = \frac{1}{\sqrt{r_k T}}. \quad (2)$$

Equation (2) shows the key dependency: for a fixed underlying rate r_k , the *relative noise in counts s⁻¹* decreases as $1/\sqrt{T}$, i.e. doubling dwell time (or the number of sweeps) improves the per-channel signal-to-noise as

$$\text{SNR}_k \approx \sqrt{r_k T} \propto \sqrt{T}.$$

Since $T = N\tau$, decreasing the sweep count N linearly reduces T ; see [3] for experimental evidence and discussion. With the growing adoption of data-driven models and Physical AI, which combines machine learning with domain-specific physics [6] [4], hybrid approaches are increasingly used to reduce time and improve efficiency.

II. METHODOLOGY

Each acquired spectrum is treated as a high-dimensional feature vector comprising intensity values sampled across m discrete energy channels. For a dataset containing n spectra, the measurements are organized into the matrix

$$X = \begin{bmatrix} I_1(E_1) & \cdots & I_1(E_m) \\ I_2(E_1) & \cdots & I_2(E_m) \\ \vdots & \ddots & \vdots \\ I_n(E_1) & \cdots & I_n(E_m) \end{bmatrix}, \quad (3)$$

where $I_i(E_j)$ denotes the measured intensity at energy channel E_j for sample i . The resulting data matrix $X \in \mathbb{R}^{n \times m}$ forms the basis for subsequent processing.

To compensate for increased shot noise from reduced acquisition time (a critical constraint in high-throughput industrial settings), the spectra undergo processing through an unsupervised machine learning-based denoising framework. The approach exploits the intrinsic structure of spectroscopic data: physically meaningful features such as peak positions, line-shapes, and relative intensities exhibit strong inter-correlation and coherent patterns, whereas random noise contributions are uncorrelated and distributed across the measurement domain. The algorithm discovers underlying statistical regularities without requiring labeled training examples, learning to distinguish between coherent signal patterns and high-frequency stochastic fluctuations by analyzing correlation structures across multiple spectra and energy channels. The denoised spectra $\tilde{X}$ preserve quantitatively accurate peak intensities

while suppressing noise components. Model parameters are optimized through validation to ensure physically plausible spectral characteristics.

Figure 2 provides an overview of the complete processing pipeline from raw measurements through denoising to quantitative concentration determination.

Machine Learning – Overview

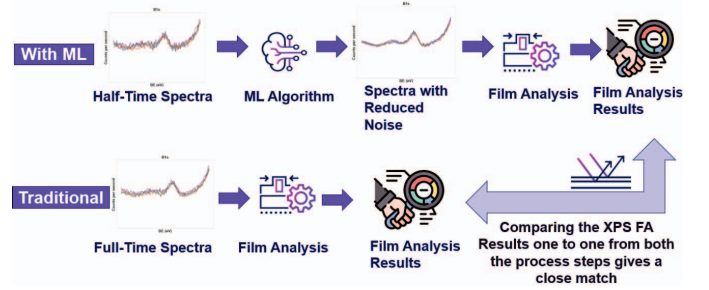


Fig. 2: Overview of ML application.

Elemental concentrations are determined from the enhanced spectra using an established thin-film analysis algorithm that accounts for multi-layer structures and incorporates standard corrections. The fundamental relationship can be conceptually represented as:

$$C_i = \frac{I_i/S_i}{\sum_j I_j/S_j} \times 100\%, \quad \sum_i C_i = 100\%, \quad (4)$$

where I_i represents the characteristic intensity for element i and S_i denotes the relative sensitivity factor. This expression provides a simplified conceptual framework; the actual implementation is substantially more elaborate and incorporates standard parameters and correction procedures appropriate for rigorous thin-film quantification. For the quaternary Si-O-B-Ge systems examined in this work, the four elemental concentrations sum to 100%. The integration of machine learning-based denoising with established film analysis enables reliable concentration determination from significantly noisier input data than conventional approaches, facilitating substantial reductions in per-spectrum acquisition time while maintaining analytical precision.

III. EXPERIMENTAL SETUP

As shown in Figure 3 the setup of X-ray photoelectron spectroscopy (XPS) quantification of boron in epitaxial SiGe source/drain structures. It is intrinsically challenging due to multiple compounding factors. The low B1s photoemission yield, combined with the low boron dose employed in the PFET epitaxial process, results in a weak boron signal. The B1s peak lies directly on top of the second Si2s plasmon, causing the boron signal to be effectively masked by the silicon plasmon background. This spectral overlap, together with three-dimensional FinFET geometry effects and matrix- and strain-induced peak broadening in SiGe, leads to a severely reduced signal-to-noise ratio (SNR). Consequently,

reliable boron quantification using conventional XPS analysis requires long acquisition times. These limitations motivate the investigation of machine learning-based approaches to reduce acquisition time while maintaining measurement accuracy and physical consistency.

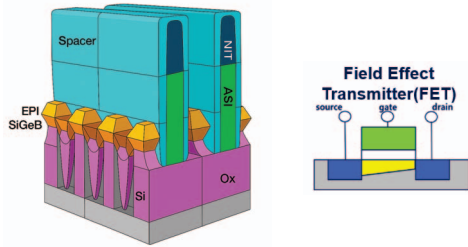


Fig. 3: Representative FinFET 3D structure with diamond-shaped SiGeB epitaxy. Derived from: Jusang Lee, Proc. of SPIE Vol 10959, 2019 [5].

Two sets of Silicon-Germanium-Boron (SiGeB) wafers (30 wafers each) were characterized using Nova's VeraFlex XPS metrology tool on regular production wafers. We used two different modes of validation. Shadow Mode and Solo Mode. Figure 4 illustrates the process flow of validation (Both Shadow and Solo mode).

- Shadow Mode:** In the Shadow Mode phase, the machine learning (ML) framework was developed and evaluated using synthetically generated half-time (HT) data derived from a first set of legacy Process-of-Record (POR) full-time (FT) acquisitions. A dedicated signal-processing algorithm was developed to generate HT-equivalent data from the original FT measurements. The algorithm was independently validated in separate experiments and demonstrated to accurately reproduce the characteristics of real half-time (HT) acquisitions while preserving the same physical wafer population. This ensured that both HT (ML-applied) and FT (reference) results corresponded to identical wafers, thereby eliminating wafer-to-wafer variability and isolating the effect of acquisition time reduction.

The ML model was trained on historical FT data and subsequently validated on the generated HT-equivalent dataset. The predicted metrology outputs were directly benchmarked against the original FT POR measurements. As shown in Fig. 4, Result 1 (POR FT baseline) is compared with Result 2 (HT-equivalent data with ML applied). This controlled evaluation enabled quantitative assessment of prediction accuracy, systematic bias, and variability under reduced acquisition conditions.

Upon meeting predefined performance criteria, the trained model was advanced to validation under real reduced-time acquisition conditions. In this step, the recipe time was physically reduced to true half-time operation (approximately 50% of the original POR FT duration), transitioning from synthetic evaluation to production-level validation.

- Solo Mode:** In Solo Mode, a second and independent wafer set was measured using the physically reduced acquisition time (approximately 50% of the POR FT recipe). Unlike Shadow Mode, measurements were collected directly under real half-time hardware conditions. The ML model developed during Shadow Mode was applied without retraining to these HT measurements. Performance was evaluated relative to the legacy POR FT baseline to verify that throughput improvement from reduced acquisition time did not compromise metrology accuracy or process control capability. As shown in Fig. 4, Result 3 (POR FT reference) is compared against Result 4 (real HT + ML). Together, Shadow and Solo Modes constitute a two-stage validation methodology: first isolating ML reconstruction capability under controlled conditions, and subsequently confirming robustness under true production operation.

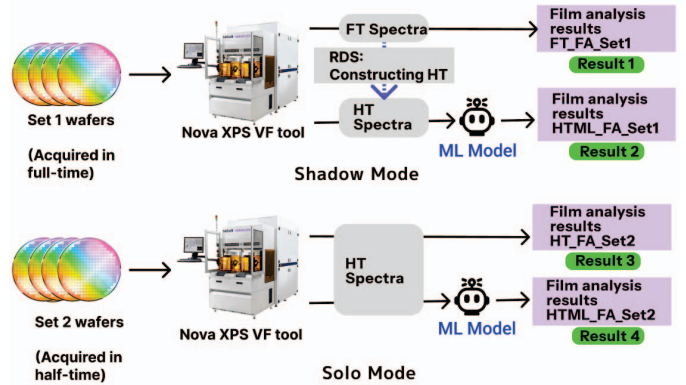


Fig. 4: Operational modes for data acquisition: Shadow Mode and Solo Mode.

Measurement sites: Thirteen (13) measurement sites per wafer capture within-wafer variation. Figure 5 illustrates the site layout used in production.

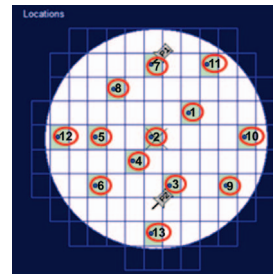


Fig. 5: Measurement site locations at each wafer

Here, “recipe” denotes the VeraFlex parameter set for SiGeB (X-ray line, analyzer pass energy, lens mode, step size ΔE , neutralizer settings).

IV. METRICS AND ACCEPTANCE CRITERIA

We compare HTML (HT+ML) to FT and contrast with uncorrected HT using HVM-oriented metrics.

1) RMSE of wafer means: Let $C_i^{(1)}$ and $C_i^{(2)}$ denote wafer-mean concentrations from the reference and candidate results, respectively.

The RMSE is

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (C_i^{(2)} - C_i^{(1)})^2} \quad (5)$$

Acceptance: $\text{RMSE} < 0.45$

2) Slope of wafer-mean line: From the linear regression $C_i^{(2)} = m C_i^{(1)} + b$, require

$$0.9 < m < 1.0. \quad (6)$$

3) Relative RSD Difference (stability): For wafer i and result k , $RSD_i^{(k)} = \sigma_i^{(k)} / \mu_i^{(k)}$. Define

$$\Delta_i = 100 \cdot \frac{RSD_i^{(2)} - RSD_i^{(1)}}{RSD_i^{(1)}} \Rightarrow \Delta_{\text{avg}} = \frac{1}{n} \sum_{i=1}^n \Delta_i, \quad (7)$$

with acceptance $-5 < \Delta_{\text{avg}} < 0$ (%).

4) Site-Scaled Standard Deviation Ratio (Cross-Wafer Stability): To evaluate cross-wafer stability, each site measurement is first normalized by its corresponding wafer mean:

$$r_{ij}^{(k)} = \frac{c_{ij}^{(k)}}{\mu_i^{(k)}},$$

where $c_{ij}^{(k)}$ denotes the measurement at wafer i , site j , for condition k , and $\mu_i^{(k)}$ is the mean across all sites on wafer i .

For each site j , the cross-wafer standard deviation is computed as

$$\sigma_j^{(k)} = \text{StdDev} \left(r_{1j}^{(k)}, \dots, r_{nj}^{(k)} \right).$$

The site-averaged standard deviation is then defined as

$$\bar{\sigma}^{(k)} = \frac{1}{13} \sum_{j=1}^{13} \sigma_j^{(k)}.$$

The stability ratio between the two conditions is given by

$$\text{Ratio} = \frac{\bar{\sigma}^{(2)}}{\bar{\sigma}^{(1)}}, \quad (8)$$

with an acceptance criterion of $\text{Ratio} \in [0.90, 1.00]$. This metric quantifies the relative change in measurement precision across wafers.

5) Point-level R^2 (all sites, all wafers): Regress $y = mx + b$ at the point level with $x_{ij} = c_{ij}^{(1)}$ and $y_{ij} = c_{ij}^{(2)}$; require

$$0.90 < R^2 < 1.0. \quad (9)$$

A/B statistical tests: Two-sample t-tests (significance $\alpha = 0.05$) compare *FT* vs. *HT* and *FT* vs. *HTML* for Boron and Silicon to assess population similarity (fail to reject H_0 indicates no statistically significant difference).

V. RESULTS

In this work, FT refers to the measurement obtained with full-time acquisition, while HT denotes half-time acquisition. HTML represents the results obtained by applying machine learning to half-time acquisition. For validation, two batches of the process are used: Shadow mode and Solo mode. HTML measurements on top of the ML model demonstrate quantitative equivalence to FT POR results, with strong linear correlation ($R^2 = 0.96$ for boron and $R^2 = 0.95$ for silicon) and slopes within acceptance limits. Figure 6 shows the accuracy plots for both boron and silicon species, confirming this strong linear agreement. Although a bias term of $b = 1.44$ for Boron and $b = 1.89$ for Silicon is present, it does not materially affect quantitative interpretation of the measurements. The bias introduces a uniform proportional offset across all values rather than altering relative trends, spatial non-uniformities, or wafer-to-wafer variations. As a result, comparative analysis and process monitoring remain unaffected. In practical applications, this bias can be compensated through calibration, and therefore it is not expected to impact the reliability or usability of the measurement results. Precision analysis shows site-level variability, and in both shadow mode and solo mode, the HTML performance are statistically indistinguishable from FT, while uncorrected HT data exhibit significant shift. Figure 7, 8, 11 and 12 shows precision plots for both shadow and solo mode validation. The sensitivity analysis indicates that, in Shadow Mode, ML-corrected HT measurements (HTML) closely match the FT baseline across wafer sites for silicon, while uncorrected HT data show elevated site-level variability. For boron, however, the reduced acquisition time introduces some loss in sensitivity that is only partially recovered by the ML correction. Figure 9 and 10 shows sensitivity plots for shadow mode validation. A two-sample t-test at the 95% confidence level confirms no detectable difference between HTML and FT, establishing that the ML correction preserves both accuracy and precision, thereby enabling throughput enhancement without compromising metrology fidelity.

Shadow Mode Precision plots for Boron and Silicon

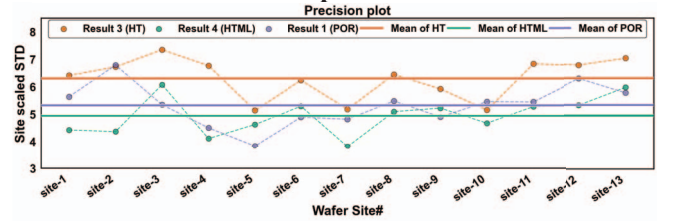


Fig. 7: Shadow Mode - Precision plot for Boron (Atomic Conc.)

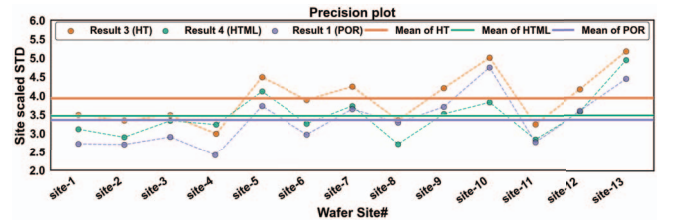


Fig. 8: Shadow Mode - Precision plot for Silicon (Atomic Conc.)

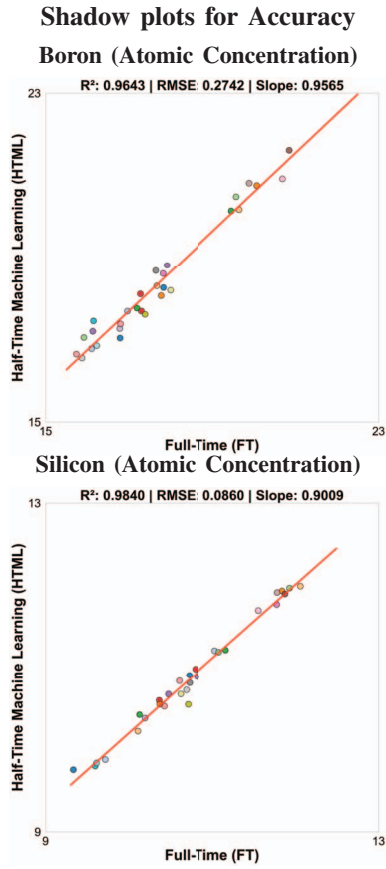


Fig. 6: Shadow Mode – Accuracy plots for Boron and Silicon.

Shadow Mode Sensitivity plots for Boron and Silicon

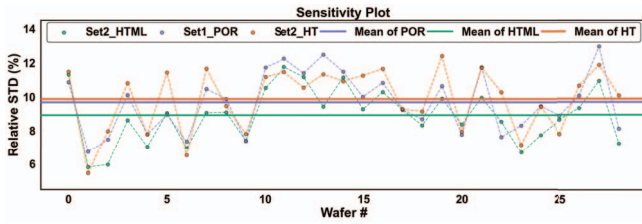


Fig. 9: Shadow Mode - Sensitivity plot for Boron (Atomic Conc.)

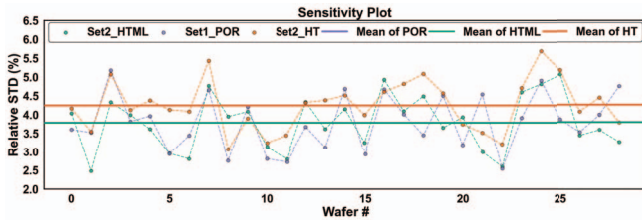


Fig. 10: Shadow Mode - Sensitivity plot for Silicon (Atomic Conc.)

Solo Mode Precision plots for Boron and Silicon

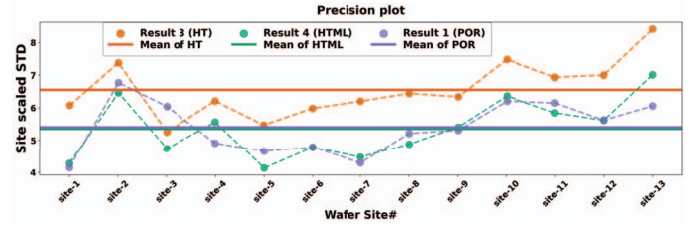


Fig. 11: Solo Mode - Precision plot for Boron (Atomic Conc.)

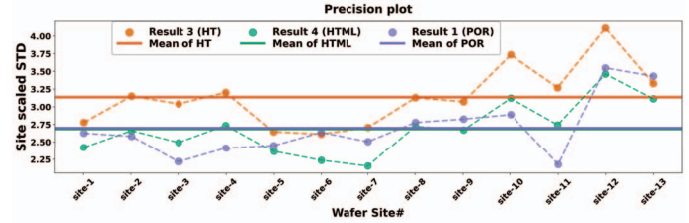


Fig. 12: Solo Mode - Precision plot for Silicon (Atomic Conc.)

Statistical A/B testing: The null hypothesis stated that the absolute difference in means between FT and the comparator exceeded an acceptable threshold ($|\mu_d| \geq \Delta$), while the alternative hypothesis stated that the difference remained within this threshold ($|\mu_d| < \Delta$). A negligible threshold of $\Delta \approx 0.2\%$ in atomic concentration was defined for silicon, with the corresponding boron threshold obtained by proportional scaling.

Let $\mu_d = \mu_{FT} - \mu_{HTML/HT}$ denote the mean difference. The hypotheses were defined as

$$H_0 : |\mu_d| \geq \Delta \quad \text{vs.} \quad H_1 : |\mu_d| < \Delta.$$

TABLE I

Statistical Test Results

Comparison	Boron	Silicon	Reject H_0
FT-HTML	0.02710	0.00102	True
FT-HT	0.63780	0.61723	False

Statistically significant p-values for the FT–HTML comparison led to rejection of the null hypothesis for both boron and silicon, supporting agreement between FT and HTML. In contrast, non-significant p-values for the FT–HT comparison indicated that the null hypothesis could not be rejected for either element.

SUMMARY

In this work, we present a production-oriented, AI-assisted spectral reconstruction framework for X-ray Photoelectron Spectroscopy (XPS) that substantially accelerates compositional metrology while preserving full-time (FT) measurement fidelity. The approach was rigorously validated on two independent SiGeB wafer sets, each comprising approximately 60 wafers, and evaluated across two complementary validation modes.

Across both datasets, the method consistently delivers FT-grade performance from half-time (HT) acquisitions, effectively doubling throughput. Key metrology metrics meet or

exceed acceptance thresholds, demonstrating robustness across spatial sampling scales. A/B testing establishes statistical equivalence for Boron and near-equivalence for Silicon relative to FT measurements. A modest reduction in Boron sensitivity was observed; however, the deviation remains within process control and manufacturing specification limits, and ongoing region-specific optimization is expected to close this gap.

This capability addresses a long-standing throughput bottleneck in semiconductor manufacturing, where XPS dwell times constrain tool utilization. By nearly doubling throughput without hardware modifications or meaningful accuracy loss, the method enables more efficient process monitoring and scales naturally within existing XPS toolchains. If HT RSTD exceeds HTML RSTD on our live monitoring dashboard, an alert is triggered to indicate potential model drift. This alert signals that retraining or recalibration of the solution may be required.

Future work will focus on extending the proposed framework across multiple XPS tool types, product lines, and process chemistries. In parallel, standardized operating procedures (SOPs) are being developed to support consistent deployment across tools and fabs. The proposed framework is structured to support applicability across a broad range of operational conditions while maintaining stable and predictable performance. Consistent evaluation and validation practices are employed to ensure reliable behavior in diverse measurement contexts. Empirical studies demonstrate stable performance under representative conditions, enabling dependable use for comparative analysis and ongoing monitoring. To support long-term robustness, the framework accommodates periodic incorporation of newly available data as part of routine operation. In addition, monitoring of measurement consistency over time is supported to ensure sustained performance as operating conditions evolve. Collectively, these characteristics provide a scalable and maintainable foundation for applying data-driven methods to advanced semiconductor metrology.

ACKNOWLEDGMENT

We thank Tyler Lynn and Robert Johnson (GlobalFoundries) for setting up the real-time NovaFit server and operational support during in-fab evaluation.

The authors gratefully acknowledge the late Alok Vaid for his pioneering contributions to the field and his invaluable guidance during the early stages of this work. His vision and expertise continue to inspire our research efforts.

REFERENCES

- [1] Neal Fairley, Pascal Bargiela, Adam J. Roberts, Vincent Fernandez, and Jonas Baltrusaitis. Practical guide to understanding goodness-of-fit metrics used in chemical state modeling of x-ray photoelectron spectroscopy data by synthetic line shapes using nylon as an example. *Journal of Vacuum Science & Technology A*, 41(1):013203, 2023.
- [2] Gordon R. Gilmore. *Practical Gamma-Ray Spectrometry*. John Wiley & Sons, 2 edition, 2008. See Ch. 5, “Statistics of Counting”.
- [3] Grzegorz Greczynski and Lars Hultman. A step-by-step guide to perform x-ray photoelectron spectroscopy. *Journal of Applied Physics*, 132(1):011101, 07 2022.
- [4] Parikshit Jain, Shane McQuarrie, and Boris Kramer. *Performance comparison of data-driven reduced models for a single-injector combustion process*.
- [5] Jusang Lee, Ganesh Subramanian, Manasa Medikonda, Hossam Lazkani, Judson Holt, Churamani Gaire, Paul Isbester, and Mark Klare. Full structure transistor process monitoring of boron and germanium in pfet epi using in-line xps. In *Metrology, Inspection, and Process Control for Microlithography XXXIII*, volume 10959 of *Proceedings of SPIE*, page 109590N, San Jose, California, USA, 2019. SPIE.
- [6] Renee Swischuk, Boris Kramer, Cheng Huang, and Karen Willcox. Learning physics-based reduced-order models for a single-injector combustion process. *AIAA Journal*, 58(6):2658–2672, June 2020.
- [7] John F. Watts and John Wolstenholme. *An Introduction to Surface Analysis by XPS and AES*. John Wiley & Sons, 2 edition, 2020.