Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/XXXX.2022.1234567

Hyperbolic Deep Learning System for Intelligent Gate-Driving and Health Monitoring of Silicon-Carbide Power MOSFETs

Francesco Rundo[1], Angelo A. Messina[2], Michele Fiore[2], Michele Calabretta[2], Pasquale Coscia[3] and Sebastiano Battiato[1], Senior Member, IEEE

[1]Department of Mathematics and Computer Science, University of Catania, Italy
[2]STMicroelectronics, QMT R&D Division, Catania, Italy
[3]Department of Computer Science, University of Milan,Italy

Corresponding author: Francesco Rundo (email: francesco.rundo@unict.it).

This work was supported by the European Union's Horizon Europe research and innovation program under the projects "R-PODID" (HORIZON-KDT-JU-2021-2-RIA-Focus-Topic-1, Grant Agreement No. 101097300) and "ARCHIMEDES" (HORIZON-KDT-JU-2021-2-RIA-Focus-Topic-1, Grant Agreement No. 101112295)

ABSTRACT Accurate monitoring of the gate–oxide health of Silicon-Carbide (SiC) MOSFET devices is pivotal for condition-based maintenance of electrified vehicles, high-frequency converters and consumer electronic applications. A MOSFET parameter closely related to the device oxide interface is the threshold voltage. The SiC MOSFETs threshold voltage (V_{th}) drifts with interface oxide–trap accumulation and it is therefore a direct early-marker of the device degradation. The V_{th} drift monitoring is then significantly relevant in the mentioned application fields. This work proposes a non-intrusive, single-shot intelligent pipeline that infers V_{th} solely from an input data-snapshot of the steady-state MOSFET static behavior based on current-to-voltage $I-V$ family curves sampled during routine gate driving. We introduce HyperDeep, a novel Hyperbolic Möbius Deep architecture that embeds each $I-V$ input snapshot in the hyperbolic Poincaré ball, processes it through ad-hoc designed residual Möbius layers followed by hyperbolic self-attention blocks. An adaptive exponential Lipschitz regularization block compensates process-spread-induced statistical drift in SiC devices, preserving mapping stability across SiC MOSFET's wide threshold-voltage window. The proposed HyperDeep system was trained on large dataset that includes both synthetic and real data obtained through Dynamic Gate Stress (DGS) methodology applied to SiC MOSFET GEN3 devices delivered by STMicroelectronics. The proposed deep pipeline attains in average of 0.015 MSE (Mean Squared Error) in retrieving actual V_{th}, outperforming compared state-of-the-art architectures.

INDEX TERMS Deep Learning, Hyperbolic Geometry, MOSFET, Residual Lifetime, Gate driving.

I. INTRODUCTION

IN the field of power electronics a Silicon-Carbide power MOSFET device is a majority-carrier, voltage-controlled switch deployed in several applications ranging from high-voltage traction inverters, Engine Control Unit (Automotive), high-temperature-resistant power engines, fast DC–DC converters (Industrial), on-board chargers (Consumer) [1]–[3]. Thus, the SiC power MOSFET plays a key and strategic role in a wide variety of applications, consequently, device efficiency and health have to be safeguarded at all times. A typical MOSFET device integrates N^+ source and drain diffusion in a P-type body under a nanometer-scale SiO_2 gate-oxide topped by a poly-silicon (metal) gate; a rising gate-to-source bias (gate-to-source voltage V_{GS}) electrostatically inverts the body surface, forming a conductive channel beneath the oxide once the critical threshold voltage V_{th} is attained [1]. This threshold voltage—the gate-to-source bias at which a conductive channel forms—sets the required gate-

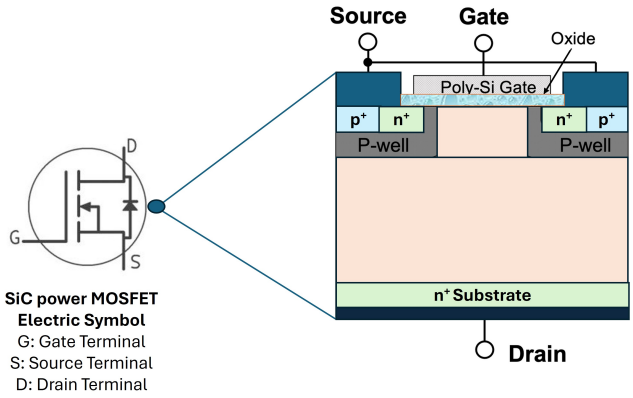

FIGURE 1. **Cross-sectional schematic and electric symbol of a vertical SiC power MOSFET, highlighting the poly-silicon gate above a thin SiO_2 oxide**

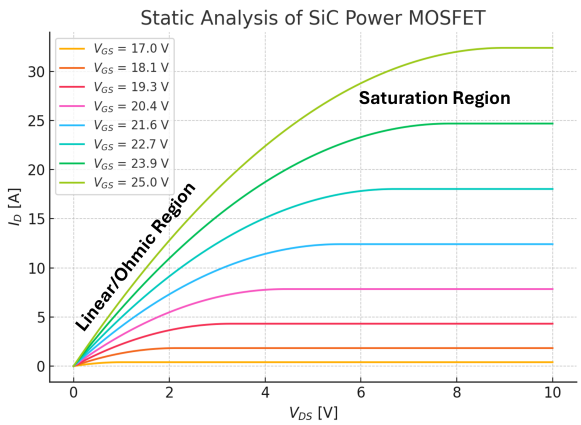

FIGURE 2. **Static I–V curves for $V_{GS} \in [17, 25]$ V; linear and saturation regions encode V_{th} and $R_{\text{DS(on)}}$.**

drive margin and is continuously monitored as a primary health and efficiency indicator [1], [2]. The threshold voltage of a power MOSFET device is not a fixed parameter: it drifts as a consequence of charge trapping in the gate oxide and related degradation mechanisms [1]–[3]. Specifically, in SiC power MOSFETs, long-term gate-bias and temperature stress inject charges into oxide/interface traps, progressively shifting the threshold voltage V_{th}; hence real-time tracking of it is needed to maintain enough gate-drive margin, diagnose aging and prevent reliability failures [1]–[4]. Fig. 1 details a typical SiC power MOSFET structure showing the mentioned poly-silicon/SiO_2 gate stack that defines the threshold voltage, the p-well surface where the inversion channel forms once this bias is exceeded, and the underlying drift region that, tied to the n^+ substrate, completes the high-voltage conduction path between source and drain terminals. In the next section, the introduced issue of the threshold voltage drift will be described.

A. SiC Power MOSFET: Threshold-Voltage Drift and Residual Useful Lifetime (RUL) Estimation

Bias–temperature instability (BTI) in SiC MOSFET causes a progressive threshold voltage V_{th} drift which has been proposed as a lifetime marker [2]–[6]. Experimental studies have documented that dynamic-avalanche conditions further accelerate this drift [2]–[4]. Real-time V_{th} tracking lets the gate driver tune V_{GS}, preserve overdrive, and curb $R_{\text{DS(on)}}$ drift [5]–[9]. Since $R_{\text{DS(on)}}$ drives conduction loss, its increase worsens self-heating and wear-out (a first-order Residual Useful Lifetime (RUL) proxy) even though online $R_{\text{DS(on)}}$ estimation is challenging [1]–[9]; V_{th} tracking is thus preferable. As V_{th} rises, overdrive shrinks and $R_{\text{DS(on)}}$ typically increases; monitoring V_{th} therefore tracks $R_{\text{DS(on)}}$ and RUL [5]–[11]. We assess inverse reconstruction of V_{th} from device I–V curves (I_D–V_{DS}) at fixed V_{GS}, which encode threshold and on-state resistance [1]–[4], [8], [9]. Fig. 2 shows the representative I–V family used by the proposed estimator. Euclidean networks often need large latent spaces for the funnel-shaped geometry (as the overall shape reported in Fig. 2); hyperbolic embeddings model it compactly, preserve hierarchies, and enable Lipschitz controls for cross-device generalization [12]. We propose a hyperbolic Möbius-attention network [12], [13] with a learnable exponential Lipschitz regularizer to mitigate MOSFET process-spread shifts [14]. Experiments confirmed the effectiveness of the proposed approach; Secs. II–VII cover principles, constraints, related work, experimental results, and conclusions regarded the proposed pipeline.

II. Hyperbolic Deep Learning: Brief Theory Recall

Hyperbolic geometry differs from its Euclidean counterpart by relaxing the parallel-line postulate, thereby endowing the space with constant negative curvature [12]. Data-points are most commonly represented in the d-dimensional Poincaré ball $\mathbb{B}^d = \{x \in \mathbb{R}^d : \|x\| < 1\}$, whose metric tensor amplifies distances near the boundary. The geodesic distance $d_{\mathbb{B}}(x, y)$ between two points $x, y \in \mathbb{B}^d$ can be computed as follow:

$$d_{\mathbb{B}}(x,y) = \operatorname{arcosh}\Big(1 + 2\,\frac{\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\Big), \qquad (1)$$

where $\|\cdot\|$ is the Euclidean norm. In hyperbolic geometry, the exponential map must be defined [12]. In this work, we introduce the exponential map $\exp_0 : T_0\mathbb{B}^d \to \mathbb{B}^d$, a function that maps a tangent vector v at the origin to the hyperbolic manifold according to the following relationship: :

$$\exp_0(v) = \tanh(\|v\|)\,\frac{v}{\|v\|}\,. \qquad (2)$$

Here, $T_0\mathbb{B}^d$ denotes the Euclidean tangent space at the origin of the d-dimensional Poincaré ball, i.e., the linear space from which vectors are lifted onto the hyperbolic manifold through $\exp_0$. The logarithmic map $\log_0 : \mathbb{B}^d \to T_0\mathbb{B}^d$ is its inverse. Because the volume of a metric ball grows exponentially with radius, hyperbolic space embeds hierarchical

data with far lower distortion than any Euclidean embedding of comparable dimension. Recent surveys [12], [13] show that hyperbolic deep neural networks have enabled several innovative intelligent blocks including Möbius linear layers, gyrovector attention and curvature-aware optimisation rules. These layers operate by shuttling activations between the manifold and its tangent spaces through the maps described in Eq. (2) so that standard deep learning training algorithm based on back-propagation can be retained with minimal overhead. The theoretical exposition in [13] further demonstrates that hyperbolic operators satisfy Lipschitz bounds analogous to their Euclidean counterparts but with tighter constants, a property that will be exploited later in this work. For the inverse retrieval of the drifted SiC MOSFET threshold voltage, the mentioned properties of hyperbolic geometry are suitable. The family of static I–V curves of a SiC power MOSFET device forms a funnel-shaped manifold whose branches diverge rapidly as the V_{DS} approaches saturation (see Fig. 2). Specifically, the branches cluster near the origin, where drain currents are uniformly low, but flare apart as V_{DS} approaches saturation, because channel conduction becomes highly V_{GS}-sensitive; hence, minute gate-bias variations produce large current differentials in the high-field region. Hyperbolic deep learning exploits the exponential volume of the Poincaré ball to linearize the funnel-shaped $I-V$ manifold, allowing the network to map tiny V_{GS} shifts to the large current swings near saturation and thereby infer V_{th} with high precision.

III. Introduction to the MOSFET Modeling

A SiC power MOSFET is a voltage-controlled, majority-carrier device whose channel forms when $V_{GS} > V_{th}$ [15], [16]. Modeling spans the *static* I–V domain and the *dynamic* switching domain [16]–[18]. Statics use two canonical families: the output $I_D(V_{DS})$ at several V_{GS}—showing a low-V_{DS} ohmic region and a saturation plateau (see Fig. 2)—and the transfer $I_D(V_{GS})$ at fixed V_{DS} for transconductance extraction [8]–[10], [15]–[18]. Dynamics analyze inductive-load commutation; typical rails are +5 to +20 V (ON-state) and -3 to -5 V (OFF-state), with V_{DS} ratings of 650–1200 V. Accurate prediction relies on SPICE-compatible compact models capturing drift resistance and channel-charge modulation [7], [15]–[19]. We propose an enhanced static SPICE Level-3 SiC model [7], [16]–[23] augmented with a physically parasitics inductance-capacitance LC "ripple" term to emulate package/circuit-layout ringing that affects the linear-to-saturation transition. Aging signatures are embedded via Dynamic Gate Stress (DGS) testing procedure [23] - JEDEC JEP192, IEC 60747-15, and ECPE AQG324 [22], [23], which accelerates degradation through repetitive gate pulses at high drain bias/temperature and captures V_{th} shift and gate-charge hysteresis [17]–[19], [23]. Given process spreads and parasitics [15]–[20], [23], the V_{th} estimator is regularized with a learnable exponential Lipschitz block. The data for V_{th} retrieval are hybrid: curves

FIGURE 3. SiC Power MOSFET in TO-247 package (SiC Gen3 1200V Power MOSFET delivered by STMicroelectronics)

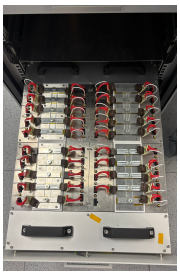

FIGURE 4. Dynamic Gate Stress (DGS) Test-Probe System Device

synthesized by the enhanced SPICE model and measured curves from DGS sessions, supplying the downstream hyperbolic system. In Fig. 3 and Fig. 4 an instance of a SiC power MOSFET (GEN3 delivered by STMicroelectronics) and DGS Probe Tesing System are reported respectively.

A. The enhanced SiC Power MOSFET Model

As described, we have collected the input dataset based on SiC power MOSFET static data by using curves generated by DGS testing sessions (applied to several SiC power devices) and data generated by an enhanced SPICE model of the SiC MOSFET [7], [15]–[23]. Preliminary, we have improved the robustness of the synthetic curve. The curves generated by this enhanced model also introduce the LC ripple phenomena introduced in the previous sections and which consist of oscillations (ripple) of the static I-V curve (Fig. 2) in the switch region from linear/ohmic region to saturation. In Fig. 5a an instance of a synthetic curve (with ripple) generated by our enhanced model is reported. In Fig. 6, an I-V curve of SiC MOSFET generated by DGS methodology with a detail of the DGS stress input waveforms, are showed.

A two-level SiC power MOSFET enhanced model is then proposed. More in detail, the drain current of the device can be represented by a two-layer formulation. The first layer reproduces the static behavior expected from a long-channel device, including temperature and channel-length effects. The second layer adds a damped sinusoidal perturbation generated by the package parasitic L (Inductance), C (Capacitance) network that becomes active once the drain voltage exceeds the saturation knee area (see Fig. 5a).

In the following Eq. (3), the threshold voltage $V_{\mathrm{th}}^{\mathrm{poly}}$ computation embedded in this enhanced SiC MOSFET static model is reported:

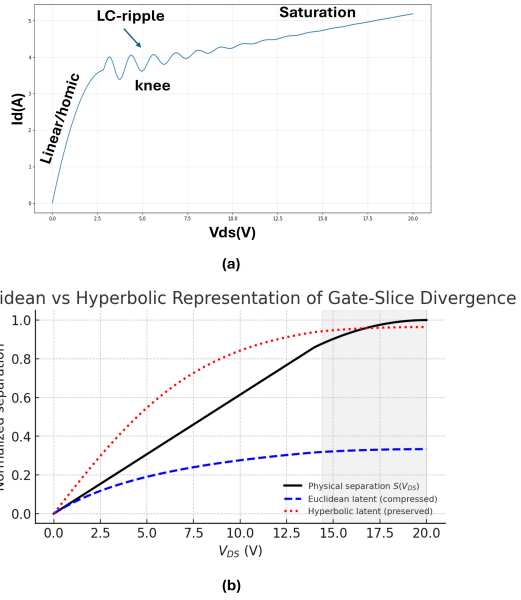

FIGURE 5. (a) Drain current versus drain–source voltage at 25°C generated by the calibrated SPICE model ("synthetic") with fitted inductance–capacitance parasitics (knee area) (b) Gate-slice separation vs. latent representation: Euclidean embeddings (blue) compress high-field differences, while hyperbolic embeddings (red) preserve the physical divergence (black).

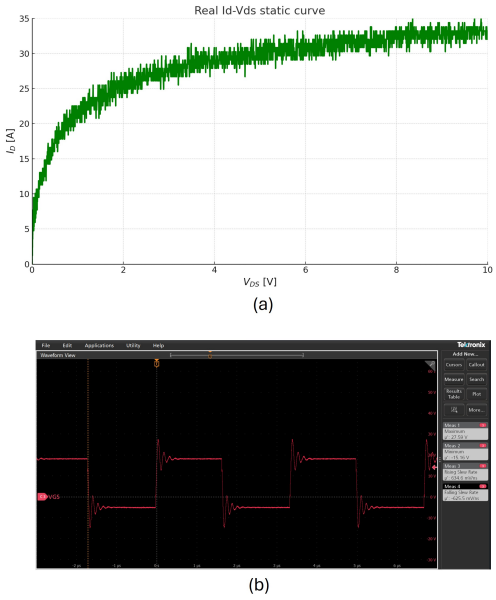

FIGURE 6. Measured static curves of SiC power MOSFET generated by the DGS procedure (at read-out window): (a) static behaviour of a Gen-3 device delivered by STMicroelectronics; (b) input-gate-stress profile used in DGS session

$$\Delta V_{\mathrm{th}}^{\mathrm{poly}} = \alpha_2 V_{\mathrm{GS}}^2 + \alpha_1 V_{\mathrm{GS}} + \alpha_0. \tag{3}$$

The quadratic expression reported in Eq. (3) captures the bias-induced threshold shift caused by poly-silicon gate charging and interface traps [1]–[8]. Coefficients α_0, α_1 and α_2 are empirical but correlate with trap density, fixed charge and poly-silicon work-function variation respectively [15]–[23]. It is known in scientific literature that the threshold voltage depends of temperature [1]–[4]. V_{th} at an arbitrary temperature T is obtained by adding the bias term $\Delta V_{\mathrm{th}}^{\mathrm{poly}}$ to the reference value V_{TO0} (used in classical SPICE Level-3 model) [19]–[23] and to a linear temperature drift. V_{TO0} is the zero-bias threshold-voltage parameter of the classical SPICE Level-3 model, i.e., the intrinsic device V_{th} measured at the reference temperature T_{REF} before any poly-charge or temperature-drift corrections are applied [1]–[4], [7], [19]–[23]. The slope $\mathrm{d}V_{\mathrm{th}}/\mathrm{d}T$ reflects the opposite trends of mobility improvement and band-gap widening in SiC devices. Finally, the enhanced mathematical modeling of V_{th} can be computed as follow:

$$V_{\mathrm{th}}(T, V_{\mathrm{GS}}) = V_{\mathrm{TO0}} + \Delta V_{\mathrm{th}}^{\mathrm{poly}} + \frac{\mathrm{d}V_{\mathrm{th}}}{\mathrm{d}T}\,(T - T_{\mathrm{REF}}). \tag{4}$$

In the classical SPICE static model, the trans-conductance parameter K_P scales the quadratic law $I_D = K_P(V_{GS} - V_{th})^2$, condensing electron mobility, gate-oxide capacitance and MOSFET channel geometry. The trans-conductance factor $K_P(T)$ depends on temperature through a power law: $K_{P0} = \mu_n C_{\mathrm{ox}} W/L$ at T_{REF}; μ_n is negative because surface mobility degrades [1]–[4], [19]–[23]. In that formula μ_n represents the electron mobility in the inversion layer, C_{ox} denotes the gate-oxide capacitance per unit-area, W and L are the effective channel width and length, respectively (see Fig. 1). The enhanced trans-conductance function can be computed as follow:

$$K_P(T) = K_{P0}\Big(\frac{T + 273.15}{T_{\mathrm{REF}} + 273.15}\Big)^{\mu_n}. \tag{5}$$

In Eq. (5) the constant 273.25 is the Celsius–to–Kelvin offset. At this point, we can proceed with computation of the V_{DS} and I_{D}. The following equation defines the drain-to-source voltage that separates the linear region from channel pinch-off [19]–[23]:

$$V_{\mathrm{DS,sat}} = V_{\mathrm{GS}} - V_{\mathrm{th}}. \tag{6}$$

The piecewise drain current definition is the long-channel square-law model reported in the following Eq. (7). The first term reflects carrier drift along the channel, while the $-\frac{1}{2}V_{\mathrm{DS}}^2$ term represents velocity reduction near the drain. In saturation, the channel current becomes independent of V_{DS} to first order. The ideal saturation drain current $I_{D,\mathrm{core}}$ model can be defined as follow:

$$I_{D,\mathrm{core}} = \begin{cases} K_P\big[(V_{\mathrm{GS}} - V_{\mathrm{th}})V_{\mathrm{DS}} - \frac{1}{2}V_{\mathrm{DS}}^2\big], & V_{\mathrm{DS}} < V_{\mathrm{DS,sat}} \\ \frac{1}{2}K_P(V_{\mathrm{GS}} - V_{\mathrm{th}})^2, & V_{\mathrm{DS}} \geq V_{\mathrm{DS,sat}} \end{cases} \tag{7}$$

Now, the LC ripple dynamic of the enhanced SiC MOSFET static model is modeled. We define the factor $1 + \xi V_{\mathrm{DS}}$ for adding channel-length modulation and drain-induced barrier lowering. Parameter ξ is inversely proportional to

the effective MOSFET channel length and approximates the finite output resistance. So, we can define the theoretical DC component of the drain current $I_{D,\mathrm{DC}}$ and the oscillatory component ω_{LC} that determines its ripple:

$$I_{D,\mathrm{DC}} = I_{D,\mathrm{core}}\big(1 + \xi V_{\mathrm{DS}}\big), \tag{8}$$

$$\omega_{LC} = \frac{F_{\mathrm{DT}}}{\sqrt{L_{\mathrm{PAR}} C_{\mathrm{PAR}}}}. \tag{9}$$

The natural oscillation frequency ω_{LC} of the MOSFET package parasitic depends on the lumped inductance L_{PAR} and capacitance C_{PAR}. Factor F_{DT} accounts for additional stray elements that slightly tune the ideal device resonance [19]–[23]. Eqs. (8)-(9) defines $I_{D,\mathrm{DC}}$ in linear/ohmic region while in saturation region, the same variable can be computed as follow:

$$I_{\mathrm{sat}} = I_{D,\mathrm{DC}}\Big|_{V_{\mathrm{DS}} = V_{\mathrm{DS,sat}}}. \tag{10}$$

The actual saturation current I_{sat} of the static model provides a convenient scaling reference for the ripple dynamic introduced previously. This LC ripple , as introduced, represents the real oscillation behavior that is embedded in the real application of the SiC power device. More in detail we define a function $s(\cdot)$ that characterize this ripple dynamic:

$$s(V_{\mathrm{GS}}) = 1 + \frac{\theta_1}{V_{\mathrm{GS}} - V_{\mathrm{th}} + \theta_2}. \tag{11}$$

Function $s(\cdot)$ increases the ripple when the gate overdrive is small, mirroring the larger impedance of the partially formed channel. The constants (θ_1, θ_2) in Eq. (11) (heuristically defined after several test-simulations) ensure numerical stability near threshold voltage. Starting from Eq. (11) we can compute the ripple amplitude:

$$A_0 = A_{\mathrm{rel}}\, s(V_{\mathrm{GS}})\, I_{\mathrm{sat}}, \tag{12}$$

where A_0 is the peak ripple magnitude. A_{rel} is a unitless coefficient that relates parasitic loop energy to the steady-state channel power and I_{sat} represents the real drain current of the device in saturation stage. After that, we can compute the function developing the typical oscillation behavior embedded in the real SiC power device especially in the switching session from ohmic to saturation region (knee area in Fig. 5a):

$$\sigma(x) = \frac{1}{1 + \exp\big(-x/\left(K_\sigma V_{\mathrm{DS,sat}}\right)\big)}. \tag{13}$$

The logistic envelope $\sigma(x)$ with $x = V_{\mathrm{DS}} - V_{\mathrm{DS,sat}}$ activates the oscillation only after the knee area of ohmic to saturation switching. K_σ coefficient controls how the ripple emerges in the switching phase. This ripple behavior produce some resistive losses:

$$d(x) = \exp\big(-\gamma x\big). \tag{14}$$

The exponential damping term in Eq. (14) models resistive losses along the current loop. Parameter γ represents the combined effect of metal resistance and magnetic core losses when external inductors are present. The excess drain voltage of the SiC device is mapped into time through weight K_T in nanoseconds per volt. This linear relation emulates the gradual phase shift of the oscillation as V_{DS} increases:

$$\tau = x\, K_T \times 10^{-9}. \tag{15}$$

Finally we can compute the full LV ripple dynamic as follow:

$$I_{\mathrm{ripple}} = A_0\, \sigma(x)\, d(x)\, \sin(\omega_{LC}\tau). \tag{16}$$

The sinusoid-based function in Eq. (16) represents the resonant current in the parasitic loop. Its amplitude is modulated by A_0, turned on by $\sigma(x)$, and attenuated by $d(x)$. The final model reported in the next Eq. (17) superimposes the ripple on the static current beyond the ohmic to saturation device switch while enforcing a non-negative drain current:

$$I_{D,\mathrm{real}} = \begin{cases} I_{D,\mathrm{DC}}, & V_{\mathrm{DS}} \le V_{\mathrm{DS,sat}} \\ \max\big[I_{D,\mathrm{DC}} + I_{\mathrm{ripple}},\, 0\big], & V_{\mathrm{DS}} > V_{\mathrm{DS,sat}} \end{cases} \tag{17}$$

This formulation enables circuit simulators to predict both the nominal static characteristic and the measurable oscillatory signature typical of fast SiC switches.

B. Hyperbolic vs. Euclidean deep learning: Motivation for Hyperbolic Latent Spaces

The underlying reasons for adopting a hyperbolic latent space is directly linked to the device physics discussed in the previous Secs. II–III. The described gate-ordered I–V family exhibits a hierarchical "funnel" structure: curves cluster tightly at low V_{DS} and diverge rapidly near saturation (see Fig. 2 for more detail), where small gate-bias variations induce large current differentials. Euclidean latent at fixed capacity tends to compress these high-field differences, leading to distortion or over-parameterization. In contrast, hyperbolic space intrinsically expands representational volume radially, faithfully preserving the divergence without degrading local resolution. Experimental comparisons in Sec. VI-A support this geometry-driven argument: the hyperbolic model consistently achieves lower error across temperature ranges (Table 2), while the Euclidean backbone underperforms (Table 4). Fig. 5b shows this effect: the physical separation between device-gate slices (black) grows steeply at high V_{DS}, yet Euclidean embeddings (blue) flatten these differences, whereas hyperbolic embeddings (red) preserve them. By matching physics-induced geometry with latent negative curvature, the hyperbolic model attains greater accuracy and robustness than its Euclidean counterpart, which suffers from curvature mismatch.

IV. RELATED WORK

Several studies in the scientific literature have investigated the MOSFET modeling and related RUL predictive markers, exploiting deep learning approaches. In [7], deep learning is leveraged for inverse modeling of semiconductor devices, including SiC MOSFETs, for threshold-voltage retrieval. In [19], the authors proposed a SPICE macro-model with interface-trap injection into threshold/mobility equations matches V_{th} drift within 3% over 25–175 °C and predicts a 12% turn-on–loss rise after gate-bias stress. In [20], a $600\,\mathrm{V}$ lateral SiC power-IC platform with Level-3 parameters predicts drain current and blocking voltage within 5% across 25–300 °C; a $0.7\,\mathrm{MHz}$ CMOS ring oscillator shows frequency drift under charge-induced V_{th} shift. The authors of [21] described a $10\,\mathrm{kV}$ 4H-SiC behavioral macro-model reproducing static/switching with $< 8\%$ error; a $20\,\mathrm{kHz}$, $370\,\mathrm{W}$ boost converter reaches 86% efficiency while turn-on loss grows with V_{th} drift. In [22], an intelligent-behavioral model for $1.5\,\mathrm{kV}$ devices fitted via PSO attains 98% static accuracy and 60% faster extraction than GA-tuning, yet still needs offline iterations to track drift. In [23] the researchers proposed a first-order Taylor surrogates for $1.2\,\mathrm{kV}$ devices (NGSpice callbacks) reducing transient-simulation time by up to 59% with switching-loss accuracy within 8%. Yang *et al.* [24] train a two-layer neural network for 4H-SiC I–V/C–V over 25–500 °C using 2% data (segmented Latin-hypercube), reducing overall error by 17.3% and linear-to-saturation error by 38.6%, but without instantaneous V_{th} retrieval. Molin *et al.* [25] measure permanent V_{th} shifts up to $0.27\,\mathrm{V}$ (8%) after $1000\,\mathrm{h}$ HTGB/gate-switching on $1.7\,\mathrm{kV}$, $45\,\mathrm{m\Omega}$ parts, projecting $\sim 20\%$ drift over ten years. Wu *et al.* [26] use a nonlinear Wiener-process with truncated-normal threshold; Bayesian updates with online $V_{\mathrm{GS(th)}}$ reduce mean RUL error from 19.45% to 9.17% ($\approx 10.3\%$ absolute). Across these works, investigaed deep learning backbones include MLP, RNN/LSTM, 1D-CNN, TCN, TFT, N-BEATS, N-HiTS, and DeepAR [7], [26]–[29]. In contrast, the proposed hyperbolic framework regresses instantaneous V_{th} from a input curves-snapshot, enabling real-time gate-bias adaptation and lifetime prediction.

V. HyperDeep System: The Proposed Hyperbolic Deep Learning System

In this section the designed hyperbolic deep system (referred as HyperDeep System) for the retrieval of the SiC power MOSFET threshold voltage and its drift is detailed. In Fig. 7 the whole proposed pipeline is reported, highlighting both training configuration of the pipeline and its inference setup suitable for final real applications. In Fig. 7 the training and inference configurations of the proposed pipeline, are reported. Training Configuration: The input is a read-out matrix $S \in \mathbb{R}^{G\times D}$ with entries $I_D(V_{DS,i}; V_{GS,j})$ acquired during a DGS switching-embedded read-out; rows index gate–source slices $\{V_{GS,j}\}$, columns sweep $\{V_{DS,i}\}$ from ohmic to saturation. DGS snapshots are augmented with synthetic curves from an enhanced SPICE model on the same (T, V_{GS}) grid; sampled dynamic DGS waveforms serve only to calibrate LC ripple of the enhanced model. Preprocessing applies the normalization in Eqs. (18)–(19), yielding a tensor X with zero-mean/unit-variance per gate slice while preserving V_{DS} ordering. A 1D-convolution Euclidean encoder along V_{DS} per slice extracts low/high-field features, projects them to the Poincaré ball; hyperbolic residual layers with attention produce context vectors, Lipschitz-regularized and a final regressor yields $\widehat{V}_{\mathrm{th}}$. Inference Configuration: The conditioning circuit acquires the same on-board snapshot $I_D(V_{DS,i}; V_{GS,j})$ and feeds the trained backbone; the health monitor converts ΔV_{th} into $R_{\mathrm{DS(on)}}$ shift and RUL via lookups derived from DGS campaigns. The proposed pipeline embeds several sub-systems which will be described in the next sections.

A. HyperDeep System in Training Configuration: The Input Sub-system

In this section the proposed "Input Sub-System" is explained. Let $\mathcal{S} = \left\{ I_D(V_{DS,i}, V_{GS,j}) \right\}_{j=1,\dots,G}^{i=1,\dots,D}$ be a $G \times D$ steady-state drain current matrix (I_D) sampled during the introduced DGS gate stress-driving merged with synthetic data generation as per model described in the previous section. Each row corresponds to a fixed gate bias $V_{GS,j}$, while each column spans V_{DS} from the ohmic region to deep saturation. To compress the six-order dynamic range inherent to power-device characteristics, the raw samples are first log-scaled:

$$X_{j,i} = \log\big(I_D(V_{DS,i}, V_{GS,j}) + \varepsilon\big), \qquad \varepsilon = 10^{-9}\ \mathrm{A}, \quad (18)$$

where ε avoids the indeterminate form at the origin without affecting large-signal values. Data normalization is then applied.

Data Normalization. Given a mini-batch $\{\mathbf{X}^{(b)}\}_{b=1}^{B}$, the sample mean μ and standard deviation σ are computed *on-the-fly* over all channels to obtain zero-mean, unit-variance tensors:

$$\tilde{\mathbf{X}}^{(b)} = \frac{\mathbf{X}^{(b)} - \mu}{\sigma + 10^{-6}}. \quad (19)$$

The small constant in the denominator prevents numerical blow-up when the low-field branch exhibits negligible variance. The collected data will be properly arranged.

Tensor organization. The so collected input snapshot is finally reshaped as $\tilde{\mathbf{X}} \in \mathbb{R}^{G\times D}$ and stacked into a three-dimensional batch tensor $\mathcal{X} \in \mathbb{R}^{B\times G\times D}$ that constitutes the input interface of the hyperbolic backbone (see next sections). The ordering (B, G, D) is chosen so that a single 1-D convolution over the V_{DS} axis can be applied independently to every gate slice, preserving the physical semantics of the measurement grid.

B. HyperDeep System in Training Configuration: The Deep Hyperbolic Sub-System

The deep hyperbolic backbone comprises five successive mappings. Preliminary, the collected normalized snapshot

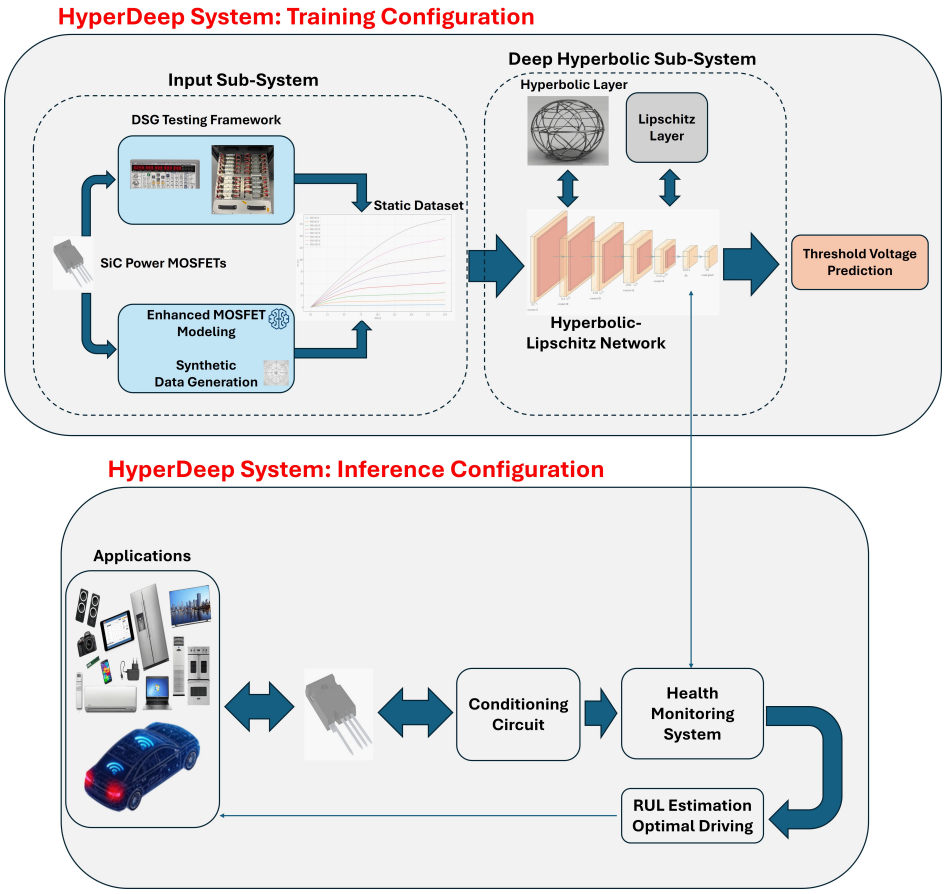

FIGURE 7. **HyperDeep: The proposed hyperbolic deep-learning system for SiC power MOSFET threshold voltage and its drift retrieval. The figure reports both training and inference configuration of the designed system**

tensor $\tilde{\mathcal{X}}$ first enters a Euclidean encoder $E_{\boldsymbol{\theta}_e}$, whose two strided one-dimensional convolutions compress each I–V slice into the latent matrix $\mathcal{E}$. This matrix is lifted into the Poincaré ball and propagated through the stack $\Phi_{\boldsymbol{\theta}_h}$ of ρ Möbius residual layers, producing the curvature-aware embeddings $\mathcal{Z}$. The embeddings are mapped to the tangent space and aggregated by the multi-head self-attention module $A_{\boldsymbol{\theta}_a}$, yielding the context vectors $\mathcal{C}$. Each context is returned to the manifold and refined by the Möbius feed-forward head (Hyperbolic MLP) $F_{\boldsymbol{\theta}_f}$, whose output manifold codes constitute $\mathcal{H}$. Finally, a Euclidean linear regressor $R_{\boldsymbol{\theta}_r}$ acting in the tangent space converts $\mathcal{H}$ into the vector $\hat{\boldsymbol{v}}_{!\text{th}}$ of predicted drifted threshold voltages. More details about each of the mentioned layers are reported in the next sub-sections.

1) Euclidean convolutional encoder

For each sample index b and gate slice j let define:

$$\mathbf{u}_{b,j}^{(1)} = \sigma_e\Big(\sum_{s=-k}^{k} \mathbf{K}_1[s]\,\tilde{\mathbf{x}}_{b,j,-s}\Big), \tag{20}$$

where $\sigma_e(t) = \max(0,t)$, $\mathbf{K}_1 \in \mathbb{R}^{c_1\times(2k+1)}$, and the argument $\tilde{\mathbf{x}}_{b,j,-s}$ is zero-padded when $i - s \notin [1, D]$. K_1 is learnable. A second convolution with weights $\mathbf{K}_2 \in \mathbb{R}^{c_2\times c_1\times(2k+1)}$ gives:

$$\mathbf{u}_{b,j}^{(2)} = \sigma_e\Big(\sum_{s=-k}^{k} \mathbf{K}_2[s]\,\mathbf{u}_{b,j,-s}^{(1)}\Big). \tag{21}$$

Averages along the drain axis yield the latent descriptor function as follow:

$$\mathbf{e}_{b,j} = \frac{1}{D/4}\sum_{i=1}^{D/4} \mathbf{u}_{b,j,i}^{(2)} \in \mathbb{R}^L, \tag{22}$$

with $L = c_2$. The Euclidean convolutional encoder compresses the raw I_D–V_{DS} data into a concise latent vector that feeds the subsequent hyperbolic regressor for threshold-voltage inference.

2) Embedding into the Poincaré ball

The above latent descriptor enters the L-dimensional Poincaré ball via the following equation:

$$\mathbf{z}_{b,j}^{(0)} = \exp_0(\mathbf{e}_{b,j}) = \tanh\big(\|\mathbf{e}_{b,j}\|\big)\frac{\mathbf{e}_{b,j}}{\|\mathbf{e}_{b,j}\|}, \qquad \|\mathbf{z}_{b,j}^{(0)}\| < 1. \tag{23}$$

The Eq. (23) provides an embedding over the Poincaré ball of the computed descriptor.

3) Möbius residual stack

Within the Poincaré ball $\mathbb{B}^L$ every transformation is built from two elementary gyrovector operators. The Möbius (gyrovector) addition of $\mathbf{x}, \mathbf{y} \in \mathbb{B}^L$ is

$$\mathbf{x} \oplus \mathbf{y} = \frac{(1 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2)\mathbf{x} + (1 - \|\mathbf{x}\|^2)\mathbf{y}}{1 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{x}\|^2 \|\mathbf{y}\|^2}, \quad (24)$$

while the Möbius scalar multiplication by any $\tau \in \mathbb{R}_+$ is computed as follow:

$$\tau \odot \mathbf{x} = \tanh\big(\tau \ \operatorname{artanh}(\|\mathbf{x}\|)\big) \frac{\mathbf{x}}{\|\mathbf{x}\|}. \quad (25)$$

Within the hyperbolic backbone we employ Möbius addition ($\oplus$) and Möbius scalar multiplication ($\odot$), which generalize linear maps to the Poincaré ball (see Eqs. 24–26). These two operations endow $\mathbb{B}^L$ with the algebraic structure required for defining linear layers, activations and residual connections that remain intrinsic to the manifold. A Möbius linear map with weights $\mathbf{W} \in \mathbb{R}^{L \times L}$ and bias $\mathbf{b} \in \mathbb{B}^L$ can be defined as follow:

$$\mathcal{M}_{\mathbf{W},\mathbf{b}}(\mathbf{z}) = \mathbf{b} \oplus \exp_0\big(\mathbf{W} \ \log_0 \mathbf{z}\big). \quad (26)$$

Define the hyperbolic SiLU activation $\sigma_{\mathbb{B}}(\mathbf{z}) = \exp_0\big(\sigma_{\mathrm{e}}(\log_0 \mathbf{z})\big)$. A single residual layer Res(z) is therefore:

$$\mathbf{r} = \sigma_{\mathbb{B}}\Big(\mathcal{M}_{\mathbf{W}_2,\mathbf{0}}\big(\sigma_{\mathbb{B}}\big(\mathcal{M}_{\mathbf{W}_1,\mathbf{0}}(\mathbf{z})\big)\big)\Big), \quad (27)$$

$$\mathrm{Res}(\mathbf{z}) = \mathbf{z} \oplus \mathbf{r}. \quad (28)$$

Iterating ρ times produces the following structure:

$$\mathbf{z}_{b,j}^{(\rho)} = \underbrace{\mathrm{Res}\big(\ldots \mathrm{Res}}_{\rho \text{ times}}(\mathbf{z}_{b,j}^{(0)}) \ldots\big), \quad (29)$$

which remains inside $\mathbb{B}^L$ and satisfies the Lipschitz inequality $d_{\mathbb{B}}(\mathbf{z}_p^{(\rho)}, \mathbf{z}_q^{(\rho)}) \leq \gamma \, d_{\mathbb{B}}(\mathbf{z}_p^{(0)}, \mathbf{z}_q^{(0)})$ with γ depending on $\|\mathbf{W}_1\|_2$ and $\|\mathbf{W}_2\|_2$. The Möbius residual stack deepens the latent representation through curvature-preserving residual transforms in the Poincaré ball, enhancing feature abstraction while retaining hyperbolic distances for accurate SiC MOSFET V_{th} regression.

4) Hyperbolic multi-head self-attention

This block is suitable to project each embedding to the tangent space, $\mathbf{T}_{b,j} = \log_0 \mathbf{z}_{b,j}^{(\rho)} \in \mathbb{R}^L$, and form classical queries, keys, and values functions [29]:

$$\mathbf{Q}_{b,j} = \mathbf{T}_{b,j}\mathbf{W}_Q, \ \mathbf{K}_{b,j} = \mathbf{T}_{b,j}\mathbf{W}_K, \ \mathbf{V}_{b,j} = \mathbf{T}_{b,j}\mathbf{W}_V,$$

with $\mathbf{W}_{Q,K,V} \in \mathbb{R}^{L \times L}$. Split into H disjoint heads of width $d = L/H$; for head h:

$$\alpha_{b,j\ell}^{(h)} = \frac{\exp\big(\langle \mathbf{q}_{b,j}^{(h)}, \mathbf{k}_{b,\ell}^{(h)} \rangle / d^{1/2}\big)}{\sum_{\ell'=1}^{G} \exp\big(\langle \mathbf{q}_{b,j}^{(h)}, \mathbf{k}_{b,\ell'}^{(h)} \rangle / d^{1/2}\big)}, \quad (30)$$

$$\mathbf{c}_b^{(h)} = \frac{1}{G} \sum_{j=1}^{G} \sum_{\ell=1}^{G} \alpha_{b,j\ell}^{(h)} \mathbf{v}_{b,\ell}^{(h)}. \quad (31)$$

Averaging the H heads gives the context $\mathbf{C}_b = \frac{1}{H} \sum_{h=1}^{H} \mathbf{c}_b^{(h)} \in \mathbb{R}^L$. The hyperbolic multi-head self-attention layer re-weights the Poincaré-embedded tokens in parallel Möbius subspaces, capturing long-range dependencies among latent features and boosting the fidelity of V_{th} estimation.

5) Möbius feed-forward head and Euclidean regressor

With this layer we embed the computed context $\mathbf{C}_b$ (Eqs. (30)-(31)) back to $\mathbb{B}^L$ and apply two Möbius linear layers with activation as above:

$$\mathbf{h}_b = \sigma_{\mathbb{B}}\big(\mathcal{M}_{\mathbf{W}_4,\mathbf{0}}\big(\sigma_{\mathbb{B}}\big(\mathcal{M}_{\mathbf{W}_3,\mathbf{0}}(\exp_0 \mathbf{C}_b)\big)\big)\big), \quad (32)$$

$$\hat{v}_{\mathrm{th},b} = \mathbf{w}_{\mathrm{r}}^{\top} \log_0 \mathbf{h}_b, \quad (33)$$

with $\mathbf{w}_{\mathrm{r}} \in \mathbb{R}^W$ and $\mathbf{h}_b \in \mathbb{B}^W$. The Möbius feed-forward head non-linearly the hyperbolic features while preserving curvature, and the downstream Euclidean regressor maps the tangent-space projection to the scalar V_{th} estimate as reported in Eq. (33). Specifically, each head from previous block is processed by two cascaded Möbius linear transformations with SiLU activation applied in the tangent space, forming a compact hyperbolic MLP that preserves negative curvature while enabling non-linear feature mixing. The per-head outputs are concatenated, projected back to the tangent space through $\log_0(\cdot)$, and passed to a single Euclidean linear layer that maps the resulting dimensional code to the scalar threshold-voltage estimate $\hat{V}_{th}$. This design blends curvature-aware representation learning with a stable Euclidean regressor, ensuring accurate and Lipschitz-controlled inference

6) Lipschitz Regularization and final Loss-Function Computation

Let $\mathbf{h}_b \in \mathbb{B}^W$ be the manifold code output by the Möbius feed-forward head for sample b and let $\mathbf{e}_b = \log_0 \mathbf{h}_b \in \mathbb{R}^W$ be its tangent embedding. To keep predictions locally Lipschitz we define the batch penalty:

$$\mathcal{P} = \frac{2}{B(B-1)} \sum_{p<q} \Big[\frac{|\hat{v}_{\mathrm{th},p} - \hat{v}_{\mathrm{th},q}|}{\|\mathbf{e}_p - \mathbf{e}_q\| + \varepsilon} - \alpha_0 \Big]_+, \quad (34)$$

where $[\cdot]_+$ is the rectifier, $\varepsilon > 0$ prevents division by zero and α_0 is the target slope. The learnable weight of $\mathcal{P}$ is $\lambda = \exp(\zeta)$ with $\zeta \in [\ln \lambda_{\min}, \ln \lambda_{\max}]$. Two data-fidelity terms complement the penalty. First, the weighted smooth-L_1 loss $\mathcal{L}_{\mathrm{H}}$ (Huber function):

$$\mathcal{L}_{\mathrm{H}} = \frac{1}{B} \sum_{b=1}^{B} \omega_b \, \psi\big(\hat{v}_{\mathrm{th},b} - v_{\mathrm{th},b}\big), \qquad \omega_b = \Big(\frac{v_{\max}}{v_{\mathrm{th},b}}\Big)^{1/2}, \quad (35)$$

with the piecewise quadratic

$$\psi(t) = \begin{cases} \frac{1}{2} t^2, & |t| \leq \beta, \\ \beta |t| - \frac{1}{2} \beta^2, & |t| > \beta. \end{cases} \quad (36)$$

Second, the mean-squared logarithmic error:

$$\mathcal{L}_{\log} = \frac{1}{B}\sum_{b=1}^{B}\big(\ln(1+\hat{v}_{\mathrm{th},b}) - \ln(1+v_{\mathrm{th},b})\big)^2. \tag{37}$$

With non-negative coefficients α_1, α_2 such that $\alpha_1+\alpha_2 = 1$, the overall training objective function is:

$$\mathcal{L}_{\mathrm{tot}} = \alpha_1\,\mathcal{L}_{\mathrm{H}} + \alpha_2\,\mathcal{L}_{\log} + \lambda\,\mathcal{P}. \tag{38}$$

Minimizing $\mathcal{L}_{\mathrm{tot}}$ over the network parameters and ζ yields a SiC power MOSFET threshold voltage predictor $\hat{v}_{\mathrm{th},b}$ both accurate and Lipschitz-controlled in the latent space.

C. HyperDeep System in Inference Configuration: The Conditioning Circuit

The conditioning circuit samples static I_{D}–V_{DS} transients at selected V_{GS} levels, providing drift-sensitive waveforms to HyperDeep for on-line V_{th} estimation, adaptive gate drive, and SiC-MOSFET RUL prediction.

D. HyperDeep System in Inference Configuration: The Heath Monitoring System

The designed HyperDeep backbone will be deployed on a resource-constrained MCU: each conditioned I–V snapshot is converted to an instant V_{th}, compared with prior values held in its Non-volatile memory, and the resulting drift is forwarded to the RUL estimator.

E. HyperDeep System in Inference Configuration: RUL Estimation/Optimal Driving

A look-up table derived in the DGS phase links ΔV_{th} to the corresponding $\Delta R_{\mathrm{DS(on)}}$. The Health-Monitoring output is therefore converted on-board into an $R_{\mathrm{DS(on)}}$ drift and, from it, a RUL estimate, while the same V_{th} value is forwarded to the gate driver for real-time optimization.

VI. EXPERIMENTAL RESULTS

In this section, the collected datasets, the designed system configuration and the benchmarking comparison results of the proposed model, will be described in detail. The study employs a dataset of 139 872 steady-state curves $I_D(V_{DS}, V_{GS})$. Real acquisitions obtained from Dynamic Gate-Stress tests supply 79 936 curves, corresponding to 57% of the total. The remaining 59 936 curves are synthetic traces generated with the temperature-aware enhanced MOSFET static model introduced previously. We have tested 9 different-types of SiC GEN3 power MOSFETs delivered by STMicroelectronics. Sample temperatures cover 25°C to 175°C in 25°C steps. During the DGS campaign every device experienced between 1.8×10^{11} and 7.0×10^{11} gate cycles at 50 % duty with $V_{GS}^{\min} = -5$ V and $V_{GS}^{\max} = 18$ V. All snapshots are log–scaled and whitened with the batch statistics before entering the network. The dataset was split device-wise into 70 % for training, 15 % for validation and 15 % for test. About data versus tested devices: As mentioned, the dataset was partitioned —five devices for training, one for validation, and three for testing—ensuring that no measurement from a training or validation device appears in the test set. In all experiments, the ripple–envelope constants were fixed at $K_\sigma = 0.08$ and $\gamma = 0.35$, values selected via a preliminary grid-search that minimized the validation MSE. Mini-batches contain $B = 256$ snapshots. The optimizer is Adam with base step size 3×10^{-4}, weight decay 10^{-4} and gradient clipping to $\|g\|_2 \le 2$. The Lipschitz weight has a separate step size 5×10^{-3}. Both rates follow a cosine schedule. Training is conducted on one NVIDIA H100-SXM-80GB GPU. The Euclidean encoder uses $c_1 = 16$ and $c_2 = 32$ channels, kernel half-width $k = 2$ and stride two. The hyperbolic latent width is $L = 192$. Four heads partition the attention layer so that $d = L/4 = 48$. The Möbius residual stack contains $\rho = 4$ layers and the feed-forward head has width $W = 48$. The Huber hinge parameter is $\beta = 1.0$ V. Data-loss weights are $\alpha_1 = 0.7$ for the Huber term and $\alpha_2 = 0.3$ for the logarithmic term. The used Lipschitz target slope is $\alpha_0 = 0.15$. The learnable penalty weight starts from $\lambda_0 = \exp(-3.9) \approx 0.019$. We set the stability term to $\varepsilon = 1\times10^{-8}$ (see Eq. (34)). Device-wise stratification (5/1/3 devices for training/validation/testing) guarantees no device leakage across splits.

The proposed pipeline has been compared with MLP, LSTM, 1D-CNN, TCN [26]–[29],Temporal Fusion Transformer (TFT), N-BEATS, N-HiTS and DeepAR architectures [29], in order to evaluate the collected performance with respect to state-of-the-art solutions. Performance is assessed on the held-out test set with Mean Squared Error (MSE). Best checkpoints were selected by lowest validation MSE under a common schedule. HyperDeep has $\approx 4.1\times10^5$ parameters; 8-bit quantization yields ≈ 0.40 MB for MCU deployment. All ablation baselines (Euclidean backbone, no-Lipschitz, no-hyperbolic) were size-matched within $\pm5\%$ and trained identically (optimizer, schedule, clipping, batch, pre-processing, device-wise splits; see Ablation Study). In Table 1, some information about the statistics of the collected static curves generated by DGS procedure are reported. Table 1 reveals a clear acceleration trend: under the baseline stress of $\pm5/18$ V at 25°C for 1.8×10^{11} cycles the average V_{th} drift stays within $3.5-5.5\%$, rises to 5.5%–6.5% at $125-150$°C with the same cycling, and peaks at 10% when both temperature (175°C) and cycle count (7×10^{11}) are maximized, confirming the synergistic impact of heat and repetitive gate stress on oxide-trap accumulation. In Table 2 the first set of benchmarks results are reported (ref. Temperature of $25-175$°C). Table 2 benchmarks mean-squared-error for V_{th} regression across 25–175 °C: The proposed HyperDeep model holds a tight 0.015 MSE and cuts error by roughly 25 % versus the next-best TFT and by about 2× compared with classical CNN, LSTM, and TCN baselines; its temperature-invariant lead underscores the benefits of hyperbolic embedding and Lipschitz regularization in mitigating thermal and process-induced variability In Fig. 8 the comparison diagram reporting the benchmark between

TABLE 1. Dynamic Gate-Stress (DGS) test matrix and observed average threshold-voltage drift

SiC MOSFET	V_{GS} range	T	Cycles	$V_{\rm th}$ drift
Test 1	$-5/18$ V	25°C	1.8×10^{11}	5.0 %
Test 2	$-5/18$ V	125°C	1.8×10^{11}	5.5 %
Test 3	$-5/18$ V	150°C	1.8×10^{11}	6.0 %
Test 4	$-5/18$ V	25°C	1.8×10^{11}	5.5 %
Test 5	$-5/18$ V	25°C	1.8×10^{11}	5.0 %
Test 6	$-5/18$ V	175°C	7.0×10^{11}	10.0 %
Test 7	$-5/18$ V	25°C	1.8×10^{11}	3.5 %
Test 8	$0/18$ V	25°C	1.8×10^{11}	6.5 %
Test 9	$-5/18$ V	25°C	1.8×10^{11}	3.5 %

TABLE 2. V_{th} drift regression: benchmark comparison at different temperatures (MSE ↓)

Architecture	25 °C	125 °C	150 °C	175 °C
MLP	0.0225	0.0228	0.0220	0.0221
1D-CNN	0.0226	0.0227	0.0224	0.0239
TCN	0.0235	0.0231	0.0235	0.0232
LSTM	0.0239	0.0240	0.0238	0.0235
TFT	0.0201	0.0201	0.0198	0.0205
N-BEATS	0.0211	0.0219	0.0219	0.0215
N-HiTS	0.0221	0.0222	0.0228	0.0221
DeepAR	0.0229	0.0227	0.0229	0.0231
HyperDeep (ours)	**0.0153**	**0.0149**	**0.0150**	**0.0152**

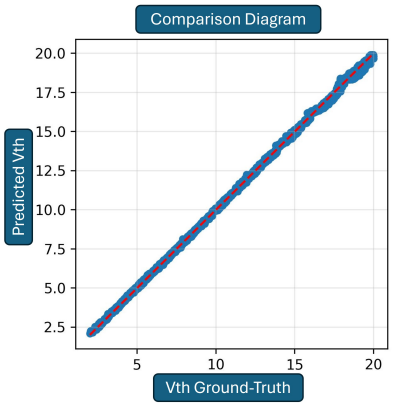

FIGURE 8. The Predicted $V_{\rm th}$ Versus $V_{\rm th}$ Ground-Truth comparison diagram

SiC power MOSFET threshold voltage ground-truth and the predicted ones, is reported. Moreover, the Lipschitz λ factor dynamic (see Eqs. (34)-(38)) embedded in the hyperbolic deep system is reported in Fig. 9.

A. Dynamic Gate Stress and Enhanced SPICE Model Calibration

As introduced, the DGS testing procedure is employed to evaluate the impact of SiC MOSFET gate oxide traps during repetitive switching, where charge trapping and de-trapping

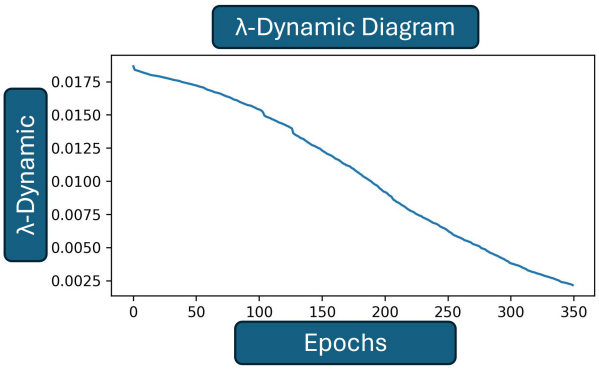

FIGURE 9. The Lipschitz regularization curve: The λ dynamic

induce measurable drifts in the gate threshold voltage $V_{\rm th}$. This stress effect is strongly influenced by the applied gate voltage range ($V_{GS}^{\rm min}$; $V_{GS}^{\rm max}$) and by LC ripple generated from parasitic inductance and capacitance in the device, package, and DGS testing-layout, with amplitude dependent on switching frequency. To capture these effects, a dynamic-enhanced SPICE model has been proposed, embedding both the static device core and package/circuit-layout induced LC ripple. The model generates synthetic I_D–V_{DS} slices and reproduces measured DGS waveforms, enabling hybrid training datasets. The performed experimental results showed that training solely on measured DGS data increases test MSE to 0.0172–0.0177 across 25–175 °C, corresponding to a 14–16.5% degradation versus hybrid datasets including synthetic curves (Table 2). Anyway, the proposed enhanced SPICE model needs parameters calibration. Model calibration was performed in two stages. Stage A (static core) fits $\{VTO_0, dV_{th}/dT, \alpha_0, \alpha_1, \alpha_2, K_P(T), \xi\}$ using static DGS read-out sweeps at multiple (T, V_{GS}), minimizing normalized RMSE of $I_D(V_{DS}; V_{GS})$. Stage B (parasitic LC) tunes $\{L_{\rm PAR}, C_{\rm PAR}, F_{DT}, \gamma, K_\sigma, K_T, \theta_1, \theta_2, A_{\rm rel}\}$ based on dynamic DGS waveform features (knee frequency, overshoot, settling), minimizing a composite cost including static RMSE, frequency error, overshoot error, and time-domain RMSE. The optimization employs Levenberg–Marquardt (LM) for Stage A and global search with LM refinement for Stage B.

B. Ablation Study

In this section an ablation analysis of the proposed pipeline is performed. As reported in Table 3, we have preliminary disabled the designed Lipschitz regularization. As confirmed in that Table 3, the performance of the proposed pipeline without the designed Lipschitz regularization is dropped significantly as the deep network was not able to compensate the statistical drift embedded in the input data coming from different SiC power devices. Finally we have removed the hyperbolic processing from the network which basically became similar to 1D-CNN in Euclidian space. In fact, the performance of the network is close to 1D-CNN compared

TABLE 3. V_{th} **regression: benchmark comparison w/o Lipschitz Regularization (MSE ↓)**

Architecture	25 °C	125 °C	150 °C	175 °C
MLP	0.0225	0.0228	0.0220	0.0221
1D-CNN	0.0226	0.0227	0.0224	0.0239
TCN	0.0235	0.0231	0.0235	0.0232
LSTM	0.0239	0.0240	0.0238	0.0235
TFT	0.0201	0.0201	**0.0198**	0.0205
N-BEATS	0.0211	0.0219	0.0219	0.0215
N-HiTS	0.0221	0.0222	0.0228	0.0221
DeepAR	0.0229	0.0227	0.0229	0.0231
HyperDeep (ours)	**0.0184**	**0.0182**	**0.0198**	**0.0197**

TABLE 4. V_{th} **regression: benchmark comparison w/o Hyperbolic Processing (Euclidean framework) (MSE ↓)**

Architecture	25 °C	125 °C	150 °C	175 °C
MLP	0.0225	0.0228	0.0220	0.0221
1D-CNN	0.0226	0.0227	0.0224	0.0239
TCN	0.0235	0.0231	0.0235	0.0232
LSTM	0.0239	0.0240	0.0238	0.0235
TFT	**0.0201**	**0.0201**	**0.0198**	**0.0205**
N-BEATS	0.0211	0.0219	0.0219	0.0215
N-HiTS	0.0221	0.0222	0.0228	0.0221
DeepAR	0.0229	0.0227	0.0229	0.0231
HyperDeep (ours)	0.0224	0.0220	0.0228	0.0233

platform as reported in Table 4, confirming the key advances of the hyperbolic analytics. To explicitly assess the role of the hybrid dataset, we performed an ablation test in which the model was trained and tested using only measured DGS acquisitions, without any synthetic augmentation. The HyperDeep pipeline remained functional but exhibited an average higher MSE compared with the hybrid setting. Across individual operating points, the DGS-only variant yielded MSEs of 0.0174 at 25°C, 0.0172 at 125°C, 0.0173 at 150°C, and 0.0177 at 175°C, corresponding to relative degradations between 14% and 16.5% over the full range. This consistent drop in accuracy highlights that the hybrid augmentation is essential for preserving robustness under temperature and process variations. Complementing DGS data with physics-based synthetic curves yields a training space that more comprehensively spans the degradation manifold, improving generalization. These findings confirmed that the performance advantage of the proposed system derives from the synergy of hyperbolic embedding, Lipschitz control, and hybrid data augmentation.

C. HyperDeep MCU deployment: Preliminary Results

To assess HyperDeep's suitability for fully standalone SiC-MOSFET health monitoring, we preliminary analyze its memory and timing footprint. We have selected a resource-constrained STM32H753 MCU delivered by STMicroelectronics [29]. The STM32H753 embeds 2 MBytes of Flash and 1 MByte of RAM (192 kB TCM + 864 kB SRAM), clocked by a 480 MHz Cortex-M7 core. Targeting a 20 kHz industrial SiC inverter $T_{inv} = 50\mu s$, we run HyperDeep system once every 1000 switching periods, i.e. every 50 ms. After 8-bit quantization, the designed pipeline stores 4.1×10^5 parameters (0.40 MB) and performs $\approx 1.4\,\mathrm{M}$ multiply–accumulates (MAC) per inference. Estimated on STM32Cube.AI framework delivered by STMicroelectronics, the forward pass completes in ≈ 2.9 ms—giving a comfortable $50/2.9 \approx 17\times$ time margin. Six SiC MOSFET devices were aged until a 3.5 % rise in $R_{\mathrm{DS(on)}}$ co-tracked with the predicted V_{th} drift—triggered an early-warning flag. Table 5 shows the detection accuracy; HyperDeep outperformed significantly the compared architectures across the 25-175 °C range.

TABLE 5. **Early-warning benchmark-accuracy(%) comparison (Accuracy ↑)**

Device	Temp. (°C)	HyperDeep	MLP	TCN	TFT	DeepAR
1	25	96.3	92.1	92.1	93.9	92.7
2	25	96.1	92.1	92.1	94.0	92.5
3	25	96.4	92.6	92.0	93.4	92.2
4	125	95.8	92.8	91.9	93.8	92.2
5	125	96.1	92.1	92.0	93.4	92.0
6	175	95.9	92.5	92.1	93.4	92.3

VII. CONCLUSION AND FUTURE WORKS

The proposed hyperbolic deep pipeline furnishes real-time, single-snapshot estimation of SiC-MOSFET threshold voltage and related drift. Trained on 1.4×10^5 hybrid curves over $25-175°\mathrm{C}$, it attains a temperature-flat $\mathrm{MSE} \approx 0.015$ (Table 2) in V_{th} prediction and $\mathrm{Accuracy}(\%) \approx 96$ (Table 5) in RUL estimation, outperforms transformer-based backbones and CNN/LSTM baselines. Ablation lifts the error to ≈ 0.018 without the Lipschitz block and to ≈ 0.022 without hyperbolic layers (Tables 3–4), validating the impact of both proposed layers. Across the DGS matrix (Table 1), the model tracks V_{th} drift, enabling closed-loop gate-bias correction and early-warning RUL diagnostics due to recalled correlation between V_{th} drift with $R_{\mathrm{DS(ON)}}$ ones. As future work, we envision hybridizing our lightweight architecture with selective physics penalties—e.g., enforcing consistency with the enhanced SPICE law in the knee/saturation regions—while maintaining the strict memory and latency budgets required for MCU deployment. This would bridge our geometry/stability-informed approach with the more rigorous physics-enforcement philosophy of physics-informed neural networks (PINNs) , potentially yielding further improvements in accuracy and explainability [30], [31].

REFERENCES

[1] Ruoyin Wang, Xiaoyong Zhu, An online junction temperature detection circuit for SiC MOSFETs considering threshold voltage drift compensation, Microelectronics

Reliability, Volume 163, 2024, 115548, ISSN 0026-2714, https://doi.org/10.1016/j.microrel.2024.115548.
[2] Chaobiao Lin, Ling Hong, Ding Wu, Na Ren, Kuang Sheng, Evolution of threshold voltage in 1.2-kV planar SiC MOSFETs during repetitive UIS stressing, Solid-State Electronics, Volume 227, 2025, 109125, ISSN 0038-1101, https://doi.org/10.1016/j.sse.2025.109125.
[3] A. Deb, J. Ortiz Gonzalez, S. Jahdi, M. Taha, P.A. Mawby, O. Alatise, Modelling SiC MOSFET module threshold voltage (VTH) and impact of parallel device δVTH on short circuit robustness, Microelectronics Reliability, Volume 150, 2023, 115101, ISSN 0026-2714
[4] Chunsheng Guo, Jiapeng Li, Yamin Zhang, Hui Zhu, Meng Zhang, Shiwei Feng, Influence of temperature inhomogeneity and trap charge on current imbalance of SiC MOSFETs, Solid-State Electronics, Volume 226, 2025, 109085, ISSN 0038-1101
[5] Zhizhao Ma, Hao Su, Yuhuan Lin, Shenghua Zhou, Feichi Zhou, Xiaoguang Liu, Longyang Lin, Yida Li, Kai Chen, Comprehensive analysis of MOSFET threshold voltage extraction method considering DIBL effect from 300 K down to 10 K, Solid-State Electronics, Volume 224, 2025, 109045, ISSN 0038-1101
[6] Thomas Aichinger, Gerald Rescher, Gregor Pobegen, Threshold voltage peculiarities and bias temperature instabilities of SiC MOSFETs, Microelectronics Reliability, Volume 80, 2018, Pages 68-78, ISSN 0026-2714
[7] Massimo Orazio Spata, Sebastiano Battiato, Alessandro Ortis, Francesco Rundo, Michele Calabretta, Carmelo Pino, Angelo Messina, "Deep learning algorithm for advanced level-3 inverse-modeling of silicon-carbide power MOSFET devices," Proc. SPIE 12973, Workshop on Electronics Communication Engineering (WECE 2023), 1297309 (16 January 2024); doi: 10.1117/12.3016130.
[8] W. Gao, P. Yin, Z. Li, M. Ren, J. Zhang and B. Zhang, "Simulation and mechanism analysis of MOSFET threshold voltage drift induced by manufacturing process," 2018 IEEE International Conference on Electron Devices and Solid State Circuits (EDSSC), Shenzhen, China, 2018, pp. 1-2, doi: 10.1109/EDSSC.2018.8487079.
[9] H. Jiang, X. Zhong, G. Qiu, L. Tang, X. Qi and L. Ran, "Dynamic Gate Stress Induced Threshold Voltage Drift of Silicon Carbide MOSFET," in IEEE Electron Device Letters, vol. 41, no. 9, pp. 1284-1287, Sept. 2020, doi: 10.1109/LED.2020.3007626.
[10] X. Zhong et al., "Recovery Performance of the Dynamic Threshold Voltage Drift of Silicon Carbide MOSFETs," in IEEE Transactions on Power Electronics, vol. 39, no. 6, pp. 7620-7631, June 2024, doi: 10.1109/TPEL.2024.3377449.
[11] N. A. Othman, N. M. Karim, M. Sufyan and N. Soin, "Impact of processing parameters on low-voltage power MOSFET threshold voltage considering defect generation," RSM 2013 IEEE Regional Symposium on Micro and Nanoelectronics, Daerah Langkawi, Malaysia, 2013, pp. 231-234, doi: 10.1109/RSM.2013.6706516.
[12] Mettes, P., Ghadimi Atigh, M., Keller-Ressel, M. et al. Hyperbolic Deep Learning in Computer Vision: A Survey. Int J Comput Vis 132, 3484–3508 (2024). https://doi.org/10.1007/s11263-024-02043-5
[13] W. Peng, T. Varanka, A. Mostafa, H. Shi and G. Zhao, "Hyperbolic Deep Neural Networks: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 12, pp. 10023-10044, 1 Dec. 2022, doi: 10.1109/TPAMI.2021.3136921
[14] Kevin Scaman and Aladin Virmaux. 2018. Lipschitz regularity of deep neural networks: analysis and efficient estimation. In Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). Curran Associates Inc., Red Hook, NY, USA, 3839–3848.
[15] Lei Tang, Huaping Jiang, Jin Wei, Qiaozhen Hu, Xiaohan Zhong, Xiaowei Qi, A comparative study of SiC MOSFETs with and without integrated SBD, Microelectronics Journal, Volume 128, 2022, 105576, ISSN 1879-2391
[16] Daniel Johannesson, Muhammad Nawaz, Analytical PSpice model for SiC MOSFET based high power modules, Microelectronics Journal, Volume 53, 2016, Pages 167-176, ISSN 1879-2391,
[17] Alessandro Sitta, Giuseppe Mauromicale, Michele Fiore, Michele Calabretta, Reliability assessment of SiC power MOSFETs in dynamic reverse bias test, Microelectronics Reliability, Volume 171, 2025, 115770, ISSN 0026-2714,
[18] SiC Power MOSFET GEN3 delivered by STMIcroelecronics: https://www.st.com/resource/en/flyer/fl2303sicgen3.pdf (Accessed on June, 2025)
[19] Y. Zhou, H. Liu, T. Yang and B. Wang, "SPICE modeling of SiC MOSFET considering interface-trap influence," in CPSS Transactions on Power Electronics and Applications, vol. 3, no. 1, pp. 56-64, March 2018, doi: 10.24295/CPSSTPEA.2018.00006
[20] T. Liu et al., "SPICE Modeling and Circuit Demonstration of a SiC Power IC Technology," in IEEE Journal of the Electron Devices Society, vol. 10, pp. 129-138, 2022, doi: 10.1109/JEDS.2022.3150364
[21] J. Wang et al., "Characterization, Modeling, and Application of 10-kV SiC MOSFET," in IEEE Transactions on Electron Devices, vol. 55, no. 8, pp. 1798-1806, Aug. 2008, doi: 10.1109/TED.2008.926650
[22] Z. Zhu, Y. Zhao and W. Yan, "Behavioral Modeling of SiC MOSFET Static and Dynamic Characteristics Based on Particle Swarm Optimization Algorithm," in IEEE Access, vol. 13, pp. 66994-67005, 2025, doi: 10.1109/ACCESS.2025.3542060.
[23] P. P. Kubulus, J. D. Meinert, S. Beczkowski, A. B. Jørgensen, S. Munk-Nielsen and D. Peftitsis, "Approximate SPICE Modeling of SiC MOSFETs," in IEEE Transactions on Power Electronics, vol. 40, no. 4, pp. 5201-5211, April 2025, doi: 10.1109/TPEL.2024.3510037
[24] W. Yang et al., "A Neural Network Modeling Method With Low-Rate Sampling for Wide Temperature Range SiC MOSFETs Application," in IEEE Transactions on Electron Devices, vol. 71, no. 6, pp. 3510-3517, June 2024, doi: 10.1109/TED.2024.3389628
[25] Q. Molin, M. Kanoun, C. Raynaud, H. Morel, Measurement and analysis of SiC-MOSFET threshold voltage shift, Microelectronics Reliability, Volumes 88–90, 2018, Pages 656-660, ISSN 0026-2714,
[26] Wu, Q., Xu, B., Xiao, L., Wang, Q.: A remaining useful life prediction method of SiC MOSFET considering failure threshold uncertainty. IET Power Electron. 17, 1594–1606 (2024). https://doi.org/10.1049/pel2.12611
[27] Karkulali Pugalenthi, Hyunseok Park, Nagarajan Raghavan, Prognosis of power MOSFET resistance degradation trend using artificial neural network approach, Microelectronics Reliability, Volumes 100–101, 2019, 113467, ISSN 0026-2714,
[28] F. Deng, Y. Bi, Y. Liu and S. Yang, "Remaining Useful Life Prediction of Machinery: A New Multiscale Temporal Convolutional Network Framework," in IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-13, 2022, Art no. 2516913, doi: 10.1109/TIM.2022.3200093
[29] C. Pino et al., "Intelligent Traction Inverter in Next Generation Electric Vehicles: The Health Monitoring of Silicon-Carbide Power Modules," in IEEE Transactions on Intelligent Vehicles, vol. 8, no. 12, pp. 4734-4753, Dec. 2023, doi: 10.1109/TIV.2023.3294726
[30] Q. He, J. Shao, J. C. Vasquez and B. Wei, "Explainable Physics-informed Self-attention-based Deep Learning for IGBT Health Estimation," 2024 Energy Conversion Congress & Expo Europe (ECCE Europe), Darmstadt, Germany, 2024, pp. 1-7, doi: 10.1109/ECCEEurope62508.2024.10751958
[31] F. Rundo, C. Pino, M. Calabretta, A. Sitta, A. A. Messina, S. Coffa, System and Method for Voltage Drift Monitoring, USA Patent Pending Nr. 20250020710; January 16, 2025.