



Research paper

Integrating physics-based simulations, machine learning, and Bayesian inference for accurate detection and metrology of elongated nanoscale analytes using high-frequency capacitance spectroscopy

Ehsan Khodadadian^a, Daniele Goldoni^b , Jacopo Nicolini^b , Amirreza Khodadadian^{a,c}, Clemens Heitzinger^a, Luca Selmi^b ,*

^a Research Unit Machine Learning, Department of Computer Science, TU Wien, Vienna, Austria

^b Department of Engineering “Enzo Ferrari”, University of Modena and Reggio Emilia, Modena, Italy

^c School of Computer Science and Mathematics, Keele University, Staffordshire, United Kingdom

ARTICLE INFO

Keywords:

Data augmentation
Bayesian inversion
Machine learning
Physics-based simulation
Nanoscale metrology
High-frequency impedance spectroscopy

ABSTRACT

Elongated analytes are simple general-purpose model systems for nucleic acid strands, bacteriophages, nanoplastic fibers, nanotubes, nanorods, etc., and are characterized by numerous unknowns (e.g., material composition, length, orientation, etc.) that are difficult to measure accurately in real-time.

This paper aims to advance the state-of-the-art of nanoscale sensing and metrology for these elongated, high aspect-ratio analytes, utilizing advanced data analysis methods specially developed for high-frequency capacitance spectroscopy measurements at nanoelectrode arrays. A model-based approach is proposed, integrating: (1) advanced supervised learning algorithms trained on an extensively augmented dataset derived from accurate physics-based numerical simulations; (2) a Markov Chain Monte Carlo (MCMC) Bayesian estimation framework for the parameters extraction task. The proposed algorithm achieves substantial speed enhancements while maintaining high accuracy, even at the resolution limits of the sensor.

The test case is developed for multispectral capacitance images of 200-1000 nanometers long nanorods captured with an advanced 256×256 pixels nanocapacitor array. The proposed approach minimizes the need for time-consuming, physics-based simulations in sensor behavior prediction and Bayesian inference iterations. It is applicable to other elongated nanoscale analytes whose state is defined by many parameters. As a result, a robust and scalable solution for efficient and precise metrology of elongated analytes is established for high parallelism and high throughput nanocapacitor array sensor applications.

1. Introduction

Highly parallel and scalable micro- and nano-electronic sensing platforms in complementary-metal-oxide-semiconductor (CMOS) technologies receive increasing attention since they could open new routes toward robust and cost-effective environmental and biological sensing. Following the early report in [Widdershoven et al. \(2010\)](#), and the demonstrations in [Widdershoven et al. \(2018\)](#) and [Laborde et al. \(2015\)](#), new prototypical implementations of capacitance and impedance sensor platforms have flourished ([Abbott et al., 2022](#); [Hu et al., 2021a,b](#); [Jain et al., 2024](#); [Jung et al., 2021](#); [Kumashi et al., 2021](#); [Lai et al., 2023b,a](#); [Lee et al., 2023](#); [Lopez et al., 2018](#); [Miccoli et al., 2019](#); [Nøvik et al., 2022](#); [Tabrizi et al., 2022](#); [Widdershoven, 2025](#)). Among these, nanoelectrode arrays for capacitance spectroscopy sensors demonstrated the smallest pixel size and the largest parallelism

([Widdershoven et al., 2018](#)). However, the adoption of this technology for the detection and characterization of small analytes is still hampered by insufficient knowledge and immature methods to interpret the huge amount of complex capacitance images produced by the sensor, especially when the signal amplitude becomes comparable to the hardware detection limits (1 zeptoFarad , $zF \div 1 \text{ attoFarad}$, aF , [Carminati et al. \(2011\)](#) and [Widdershoven et al. \(2018\)](#)).

A modeling framework of proven validity to describe the signal transduction mechanisms of these platforms is the linearized, time-harmonic (alternating current, AC) formulation of the Poisson–Boltzmann and Poisson–Nernst–Planck equations (hereafter denoted AC-PNP), possibly expanded with chemical reaction terms. This model well captures the space- and frequency-dependent conduction and displacement behavior of nanoelectrode biosensors, which is necessary to simulate their frequency-dependent impedance ([Barsoukov and](#)

* Corresponding author.

E-mail address: luca.selmi@unimore.it (L. Selmi).

Macdonald, 2005; Cossettini et al., 2021; Faustmann et al., 2022). Furthermore, it enables the analysis of complex electrodynamic processes and the extraction of key analyte properties from measurements in biosensing applications. However, solving the fully coupled system of nonlinear equations for potential and ionic concentrations requires a huge number of degrees of freedom (\mathcal{N}_{dof}), often reaching hundreds of thousands of unknowns in finite element discretizations (Taghizadeh et al., 2017; Khodadadian et al., 2020a). These computations are even more demanding when accounting for frequency-dependent effects, as the system must be solved repeatedly across multiple frequencies.

Accurate physics-based simulations of the sensor's signal transduction mechanisms have been used to interpret these signals (Cossettini and Selmi, 2018; Goldoni et al., 2024a; Guillén et al., 2025; Widdershoven et al., 2018) and, combined with machine learning (ML) models, to generate reliable training data sets supporting, for instance, the metrology of spherical nanoparticles from capacitance images (Lombardo et al., 2024). However, physics-based simulations can be exceedingly time-consuming, and also require uncommon competencies within the life-sciences, and the sensor end-user groups to become of widespread usage.

The use of ML techniques is advancing rapidly in the field of biosensors, supporting data processing in domains that include, but are not limited to, microfluidic devices for lab-on-chip and organ-on-chip applications (Antonelli et al., 2024; Khodadadian et al., 2022; Mirsian et al., 2025), as well as food safety monitoring (Zhou et al., 2024). Furthermore, ML techniques have facilitated sophisticated tasks such as cell sorting and classification for optical biosensors (Asadian et al., 2024; Cui et al., 2020), the analysis of signals from point-of-care sensors (Bhaiyya et al., 2024), and data fusion (Cui et al., 2020). Using ML for the AC-PNP equations (Huang et al., 2024) also offers a valid route to overcome the challenges of the nanoscale metrology problem at a much smaller computational cost.

The *first step* of this work was then to develop a supervised ML framework using high-fidelity and precisely labeled numerical solutions of the AC-PNP model as training data (Pittino and Selmi, 2014). The ML model is thus equipped to predict the capacitance images for any combination of input parameters for elongated analytes (e.g., length, permittivity, or rotation angle), without the need to re-simulate the entire system. This approach drastically reduces computational costs, enabling predictions in seconds while maintaining the accuracy of the physics-based model. Consequently, the ML model facilitates the efficient exploration of parameter space, making it an invaluable tool for interpreting experiments, estimating the analyte parameters, and designing new, optimized capacitance sensors.

To introduce the *second step* of this work, we note that measuring in real-time properties such as length, permittivity, and rotation angle for nanoscale analytes presents substantial challenges. The determination of the analyte length often requires precise control over fabrication processes and advanced, expensive imaging techniques, both of which introduce errors. Permittivity measurements are further complicated by heterogeneities in the analyte material and variations in local environmental conditions, such as ionic strength or temperature. Estimating the analyte's rotation angle is particularly difficult because it depends on molecular orientation relative to the sensor surface, which is influenced by dynamic, nonuniform interactions at the nanoscale.

Bayesian inversion (Smith, 2013) provides a robust framework for addressing these challenges by combining experimental data with prior knowledge of the system. This method allows us to estimate parameters probabilistically, accounting for uncertainties in both the measurements and the model (Noii et al., 2022). Indeed, Bayesian algorithms have demonstrated utility in biosensing applications, such as in the data analysis for plasmonic biosensors (Feng and Kepler, 2015), the denoising of signals from glucose sensors (Camerlingo et al., 2023), and estimation of geometrical parameters in nanocapacitor array biosensors (Stadlbauer et al., 2019). By incorporating AC-PNP simulations as

the forward model, Bayesian inversion could explore wide parameter spaces and provide posterior distributions that quantify confidence in the estimated analyte properties. However, using the AC-PNP as a forward model would have an unbearable computational cost.

Therefore, despite initial efforts demonstrating the potential of using ML or Bayesian techniques to estimate, for instance, the radius of spherical micro- and nano-particles (Cossettini et al., 2019; Lombardo et al., 2024) or nanoelectrodes (Stadlbauer et al., 2019), the application of ML and Bayesian methods to nanoelectronic environmental and (bio)sensing for single-analyte parameters extraction remains largely underdeveloped. In fact, training these algorithms is challenging due to transistor variability, drifts, and electronic noise, which obscure the accurate labeling of measured data for nanoscale analytes. Additionally, the lack of, or limited parallelism in most CMOS biochips slows data acquisition for diverse analytes and conditions, making the process time-intensive (Antonelli et al., 2024; Lombardo et al., 2024). Approximate Bayesian Computation (ABC) methods (Marin et al., 2012; Sunnåker et al., 2013) offer a viable alternative for Bayesian inference when the likelihood is computationally intractable, as is often the case with complex, physics-based models like the AC-PNP framework. Since ABC bypasses the need for an explicit likelihood by using simulation-based acceptance criteria, it can be advantageous in settings where forward models are expensive to evaluate.

In this complex context, our study alternatively presents a novel, comprehensive, integrated framework that combines high-fidelity physics-based simulations and ML-driven modeling with carefully balanced data augmentation methods and advanced Bayesian estimation techniques, to enhance the performance of CMOS nanoelectrode array capacitance spectroscopy sensors.

Our approach is designed to tackle the characterization of a broad class of analytes with elongated morphologies, particularly structures with large aspect ratios featuring multiple degrees of freedom. Such analytes present unique challenges due to their intricate geometry and sensitivity to nanoscale variations. Capacitance spectroscopy is inherently sensitive to the analyte morphology, spatial positioning, physical properties, and surrounding medium. However, deciphering this information from measured capacitance images is a challenge, and we approach it with:

1. ML models trained on physics-based AC-PNP simulations to generate accurate and precisely labeled capacitance predictions for varying analyte lengths, permittivities, and orientations at low computational cost.
2. Using these predictions within a Bayesian inversion framework to estimate the analyte properties while rigorously accounting for uncertainties.

This integrated methodology not only accelerates the forward modeling process (due to the availability of a fast and differentiable ML-based forward model) but also enables robust parameter estimation on a large scale in a complete Bayesian framework, significantly advancing the capabilities of nanoelectrode-based impedance spectroscopy for parallelized detection and characterization of complex analytes.

In this context, elongated shapes are a perfect test case since they serve as ideal models for various analytes, including nanoplastics and microplastics commonly found in air (Eberhard et al., 2024; Atugoda et al., 2023), water (Lee and Chae, 2021; Yusuf et al., 2022), and soil (Sajjad et al., 2022; Atugoda et al., 2023). These pollutants often appear as rods, fibers, or fibrils in environmental samples (Lee and Chae, 2021; Tran et al., 2023), with their health risks still poorly understood (Geremia et al., 2023; Osman et al., 2023; Sangkham et al., 2022). Similarly, biological entities such as bacteriophages, certain bacteria (e.g., *Escherichia coli*), macromolecules like DeoxyriboNucleic Acid, DNA, RiboNucleic Acid, RNA, and Peptide Nucleic Acid, PNA strands, and proteins can be approximated as long, thin cylinders for impedance response calculations. This geometric simplification enables accurate

modeling for environmental and biomedical applications when dimensions become comparable to the electrode pitch (i.e., the pixel size) and the capabilities of the sensing platform are pushed to their spatial resolution limits.

The manuscript is organized as follows. Sections 2 and 3 describe the reference nanoelectrode array sensor of this work and the physics-based numerical model used to predict its response and to construct the training and validation sets for the ML model. Section 4 illustrates the data augmentation techniques and the ML model used to generate virtual measurements for the Bayesian inversion parameter estimation algorithm reported in Section 5. Next, Section 6 provides an extensive report of the speed and accuracy results achieved in identifying the analyte parameters. Section 7 concludes the paper stating the key advantages of our approach and providing indications about possible future developments.

2. The nanoelectrode array sensing platform

In this work, we use second-generation chips (Widdershoven et al., 2018) of the nanoelectrode array firstly proposed in Widdershoven et al. (2010). They have been successfully employed for nanotubes' characterization (Goldoni et al., 2024a), cell discrimination (Laborde et al., 2015; Guillén et al., 2025), nanoplastics (Goldoni et al., 2023b) and microplastics (Goldoni et al., 2024b) monitoring.

The chip, designed and fabricated with 90 nm low-power CMOS technology by NXP Semiconductors, features an array of 256×256 gold-copper, AuCu, planar nanoelectrodes at sub-micron distance (Table 1), enabling charge-based capacitance measurements at each electrode with zero DC bias in the 2–70 MHz frequency range (hence, negligible electrochemical reactions and nearly ideally polarizable electrode behavior).

The analyte detection signals are given by the sudden, frequency-dependent capacitance steps ($\Delta C(f) = C_{with-analyte} - C_{without-analyte}$) generated by the analyte upon entering the detection volume above the nanoelectrode.

The platform implements numerous hardware and software solutions to suppress or mitigate drift and noise on the capacitance readings; they are extensively described in Widdershoven et al. (2018), Laborde et al. (2015) and Lemay et al. (2016), and result in an excellent resolution of ≈ 2 aF (Goldoni et al., 2023b; Widdershoven et al., 2018). Among these: the AuCu nanoelectrodes ensuring excellent uniformity and time stability; correlated double sampling before Analog–Digital Conversion, ADC; shot noise suppression; ADCs' offsets compensation; external stabilization of the chip temperature; column-by-column calibration and compensation algorithms that ensure a flat capacitance response during spectral measurements in air and yield excellent agreement with theoretical expectations in salty solutions. Furthermore, the subtraction of $C_{without-analyte}$ compensates for systematic inter-electrode variabilities (Goldoni et al., 2023b,a). The measurement proceeds by row-wise scanning of the array; the working electrodes are those on the addressed row, while all the others serve as counter electrodes. High-frequency operation ($f > 1$ MHz) strongly mitigates the Debye screening from ions in the electrolyte, thus allowing for deeper penetration of the AC electric field into the microfluidic chamber. Fig. 1(a) (top) shows the chip with the PolyDiMethylSiloxane, PDMS, seal that defines the microfluidic chamber on top of the sensing array (bottom).

Several impedance/capacitance micro/nanoelectrode array sensors have been reported in the literature (Abbott et al., 2022; Hu et al., 2021a,b; Jain et al., 2024; Jung et al., 2021; Lai et al., 2023a,b; Lee et al., 2023; Miccoli et al., 2019; Tabrizi et al., 2022). Only a few feature an electrode area density $\geq 10^4$ #electrodes per mm^2 (Hu et al., 2021b; Lee et al., 2023; Lai et al., 2023a; Miccoli et al., 2019), and even fewer exceed 10^5 (Miccoli et al., 2019). Furthermore, only those in Jain et al. (2024), Lai et al. (2023b) and Miccoli et al. (2019) measure at $f > 1$ MHz, while only (Hu et al., 2021b; Tabrizi et al., 2022) report measurements above 100 MHz. Moreover, none of these

Table 1

Geometrical and simulations parameters of the system under study. See Fig. 1(b) for the definitions. The entries in boldface and italics are the four input features of the ML model. Parameter estimation addresses the entries in boldface.

Electrolyte composition	NaCl 100 mM
Electrolyte relative permittivity ($\epsilon_{r,el}$)	80
Image size	7×7 pixels
(pitch _x , pitch _y)	(0.6, 0.89) μm (Widdershoven et al., 2018)
Electrode radius	90 nm (Widdershoven et al., 2018)
Analyte length (L)	[200–1000] nm, (10 lin-spaced values)
In-plane rotation angle (ϕ)	[0–165] ° (19 lin-spaced values)
Analyte relative permittivity (ϵ_r)	[1–5] (9 lin-spaced values)
Analyte radius (R)	50 nm
Analyte center of mass position (d_x, d_y, d_z)	(0, 0, 40) nm
Vertical rotation angle (θ)	0 °
Analyte conductivity ($\sigma_{c,yf}$)	0 S/m
Temperature	22 °C
<i>Measurement frequencies (f)</i>	<i>2–70 MHz (10 log-spaced values)</i>

systems has electrode size smaller than $1 \mu\text{m}^2$. The sensor chip of this work, instead, has $2.5 \cdot 10^6$ #electrodes per mm^2 , measurement frequency up to several hundred MHz (Cossettini et al., 2020), electrode size $\approx 0.025 \mu\text{m}^2$, sub- μm spatial resolution, and noise floor in the aF range (Widdershoven et al., 2018). These attributes make the platform of this work a unique benchmark to test advanced algorithms for electrical characterization and metrology of nanoscale analytes.

Fig. 1(b) shows the template cylindrical dielectric analyte considered in this work, and defines its parameters; the values and ranges (Table 1) are typical for nanoplastics (Goldoni et al., 2023b), viruses (Cossettini and Selmi, 2018), and nanotubes (Goldoni et al., 2024a). We focus on analytes dispersed in a 100 mM sodium chloride, NaCl, water solution, representative of various environmental monitoring applications (e.g., nano/microplastics in waters (Goldoni et al., 2023b)), and comparable to that of physiological fluids. This is also a condition where the use of high frequency platforms is best justified as an effective mean to overcome Debye screening. Finally, Figs. 1(c) and 1(d) show a representative example of an analyte system mesh (with length $L = 1000$ nm, in-plane rotation angle $\phi = 0^\circ$, and relative permittivity $\epsilon_r = 1$, all the other parameters as in Table 1) and the resulting simulated ΔC (3×3) map, respectively. Note that the calculated ΔC is negative, consistently with expectations for purely dielectric analytes (Goldoni et al., 2024b; Pittino et al., 2015).

3. The physical modeling framework

Since our platform operates at zero (direct current, DC) bias and it has nearly ideally polarizable electrodes, the static behavior of the electrolyte/analyte system is well described by the Poisson–Boltzmann Equation (PBE), which models the equilibrium potential and ion distributions. For dynamic analysis in a time-harmonic (AC) regime, we employ the linearized Poisson–Nernst–Planck equations (AC-PNP) to study frequency-dependent drift and diffusion transport of ionic and electronic species. Together, these frameworks enable the calculation of critical parameters such as impedance, admittance, and effective capacitance, which are essential for describing the electrode response, interpreting experiments, and designing optimized nanoelectronic biosensors.

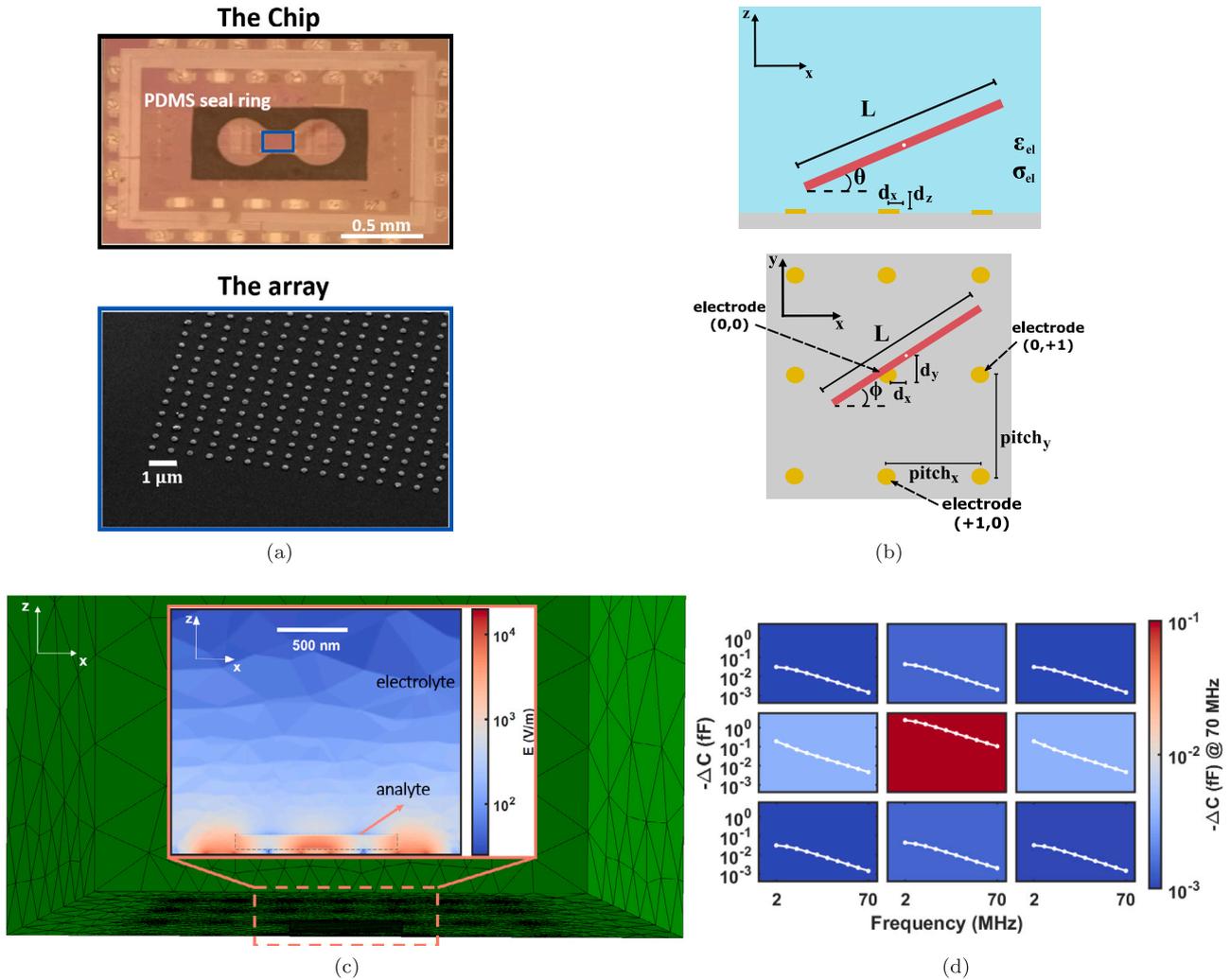


Fig. 1. (a) Sensor chip with the PDMS seal ring showing the sample flow direction over the array (top) and detail of the nanoelectrode array (bottom). (b) Sketch of the cylindrical analyte over the array with parameters' definitions. (c) 3D view of the 7×7 electrodes' domain and typical grid. The inset shows the AC electric field map above the three central (working) electrodes in the presence of the analyte. (d) 3×3 pixels zoom of the simulated hyperspectral image in log-scale of the absolute ΔC value. (c) and (d) refer to a mesh and the simulated spectra, respectively, of an analyte with length $L = 1000$ nm, in-plane rotation angle $\phi = 0^\circ$, and relative permittivity $\epsilon_r = 1$, all the other parameters are reported in Table 1.

DC formulation: Under equilibrium, zero DC bias conditions imposed by the sensor during measurements, the PBE for the DC solution reads:

$$\nabla \cdot (\epsilon(\vec{r}) \nabla V_0(\vec{r})) = -\rho_{0f}(\vec{r}) - q \sum_{m=1}^{N_{ions}} Z_m n_{0,m}^{\infty} \exp\left(-\frac{Z_m q (V_0(\vec{r}) - V_{ref})}{k_B T}\right). \quad (1)$$

Here V_0 is the DC potential, q is the elementary charge, k_B is the Boltzmann constant, ϵ is the dielectric permittivity, T is the absolute temperature, and \vec{r} is the 3D spatial coordinate. The charge density is given by immobilized charges (ρ_{0f}) and mobile ions in the electrolyte (second term). The DC concentration of the m th mobile ionic charge species is:

$$n_{0,m}(\vec{r}) = n_{0,m}^{\infty} \exp\left(-\frac{Z_m q (V_0(\vec{r}) - V_{ref})}{k_B T}\right), \quad (2)$$

The total mobile charge depends on the signed valence (Z_m) and the bulk concentration of the ionic species ($n_{0,m}$), the number of ions (N_{ions}), and the reference potential in the electrolyte bulk (V_{ref}).

AC formulation: The AC model is formulated in the time-harmonic linearized small-signal regime. Under this assumption, we obtain the

AC Poisson equation:

$$\nabla \cdot (\epsilon(\vec{r}) \nabla V(\vec{r})) + \sum_{m=1}^{N_{ions}} \frac{q_m^2 Z_m^2}{k_B T} n_{0,m}(\vec{r}) (\tilde{\phi}_m(\vec{r}) - \tilde{V}(\vec{r})) = 0, \quad (3)$$

where \tilde{V} is the electrostatic potential phasor, $\tilde{\phi}_m$ is the pseudo-potential of the m th ionic species, and $n_{0,m}$ is its DC concentration. Due to the limited amplitude of the electrode switching signal in the hardware (≈ 300 mV), the CBCM capacitance measured by the sensor is well approximated by the *effective capacitance* (Laborde et al., 2015; Pittino and Selmi, 2014; Widdershoven et al., 2018):

$$C_{\text{eff}} = \frac{|Y_A|}{\omega}, \quad (4)$$

where $Y_A = \tilde{I}/\tilde{V}_{AC}$ is the electrode admittance, \tilde{V}_{AC} is the AC potential, \tilde{I} is the AC current across the WE. The Y_A describes the dynamic response of the electrons, holes, and ions to the AC perturbation. The AC current density (\tilde{J}) is the summation of the displacement (\tilde{J}_D) and ionic current densities (\tilde{J}_m):

$$\tilde{J} = \tilde{J}_D + \sum_m^{N_{ions}} \tilde{J}_m = -j\omega\epsilon\nabla\tilde{V}(\vec{r}) - \sum_m^{N_{ions}} q^2 Z_m^2 \mu_m n_{0,m} \nabla\tilde{\phi}_m(\vec{r}), \quad (5)$$

whereas \tilde{I} is the integral of the normal component of the current density \tilde{J} over the electrode surface area, A_{el} . Assuming equilibrium conditions for the electrolyte and no DC current flow, as in the real platform, the linearized Drift-Diffusion ion transport equations take the form:

$$qZ_m\mu_m\nabla\cdot(n_{0,m}(\vec{r})\nabla\tilde{\phi}_m(\vec{r})) - j\omega n_{0,m}(\vec{r})\frac{qZ_m}{k_B T}(\tilde{\phi}_m(\vec{r}) - \tilde{V}(\vec{r})) = 0, \quad (6)$$

where μ_m is the mobility of the m th ionic species, j is the imaginary unit, and ω is the angular frequency. At low frequency ($\omega \rightarrow 0$), the AC response closely follows the DC solution, dominated by drift-diffusion processes (the first term in (6)). At high frequency, displacement currents dominate and C_{eff} reflects the frequency-dependent admittance.

Numerical simulations of the sensor response were run with ENBIOS, an in-house, accurate simulator that self-consistently solves the PBE and the coupled PNP equations for all ion species in the electrolyte. ENBIOS solves the full nonlinear system of equations using a hybrid numerical approach that integrates the control volume method (CV, which ensures strict local conservation of physical quantities, such as charge and current) and the finite element method (FEM, which uses piecewise basis functions) to achieve accurate solution approximations in complex geometries (Pittino and Selmi, 2014).

The geometries are defined and tetrahedral grids are generated using Netgen (Schoeberl, 1997). The working (WE) and counter (CE) electrodes are modeled using Dirichlet boundary conditions, while Neumann conditions are applied elsewhere. In complex 3D structures, the electrode current is calculated from the current density fluxes across the surfaces of control volumes. Further details on the ENBIOS models can be found in Pittino (2015) and Pittino and Selmi (2014).

Compared to the Scharfetter-Gummel procedure, which relies on linearization and is optimized for drift-dominated transport, our approach enables greater accuracy in systems with strong electrostatic-transport coupling. Additionally, the hybrid CV-FEM framework is better suited for handling complex geometries and non-uniform material properties, offering enhanced flexibility and applicability across both drift- and diffusion-dominated regimes. This makes it particularly effective for the precise modeling and optimization of nanoelectronic biosensors.

The effectiveness of the simulator in reproducing experimental results has been widely explored and demonstrated (Cossetti et al., 2021; Goldoni et al., 2023a,b; Laborde et al., 2015; Lombardo et al., 2024; Widderhoven et al., 2018).

To emulate the chip's parallel operating principle (Section 2), we apply \tilde{V}_{AC} to all nanoelectrodes in the central row of the array keeping at DC and AC ground all the other electrodes; then compute the admittances of all electrodes in the row as a function of the AC modulation frequency, and finally we repeat the above steps for all rows. The simulations to calculate the ΔC values according to the definition in Section 2 are first run with the analyte, and then assigning the "electrolyte" material to the analyte mesh region. Thus, the mesh is the same for the "with analyte" and "without analyte" capacitance, avoiding numerical errors due to different meshing.

By modifying the analyte length, permittivity, and orientation, many configurations are generated to be used as a-priori knowledge for the tuning of the Bayesian inversion parameter estimation algorithm. The meshes for these simulations, with L and ϕ spanning the ranges in Table 1 typically have $\mathcal{N}_{\text{dof}} = 141, 110 \pm 29, 902$ tetrahedral elements (corresponding to $26, 874 \pm 5, 091$ points) each. The calculations run on a high-performance machine with 104 Intel Xeon[®] Gold 6338 Processors (48 MB cache, 2.00 GHz) and 512 GB RAM. We used MATLAB's *Parallel Computing Toolbox* to distribute the workload across multiple cores, significantly enhancing the efficiency and speed of our simulations.

The combination of many degrees of freedom, nonlinear systems, frequency-dependent solutions, and the choice of an accurate method like CV-FEM, naturally results in high computational costs. This is

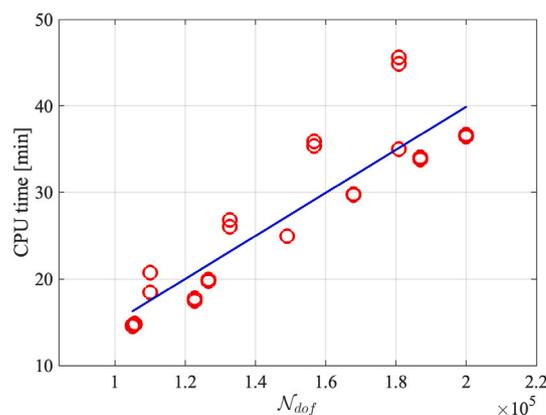


Fig. 2. PDE solver CPU time as a function of \mathcal{N}_{dof} .

illustrated in Fig. 2, reporting average CPU times in excess of 30 min per typical run. Such extensive CPU requirements pose a significant challenge for parameter estimation tasks, as these necessitate hundreds or even thousands of simulations for reliable inference. To address these computational challenges, we have developed a supervised learning framework, which is described in the following Section 4 and investigated in Sections 6.1 and 6.2. This framework, trained on a limited number of ENBIOS simulations, efficiently computes the capacitance at the sensing points.

4. The machine learning modeling framework

Given the significant computational expense of solving the AC-PNP equations during the Bayesian parameter estimation iteration outlined in Section 5, we developed an accurate ML model to predict sensor responses to analytes. The ML model was trained on a data set that encompasses a wide range of the most influential system parameters, including analyte length, permittivity, and in-plane rotation angle.

The training dataset consisted of 8,550 samples, each having four input features (L , ϕ , ϵ_r , and f) and 49 output labels (the 7×7 ΔC array). To construct this dataset, we began with AC-PNP simulations that systematically varied key parameters and applied a carefully designed data augmentation strategy incorporating small perturbations, random sampling within parameter ranges, and synthetic data generation via generative adversarial networks (GANs). These three distinct augmentation techniques contribute an additional 2,850 samples each, resulting in 8,550 augmented samples that effectively double the size of the training/testing/verification sets and enable comprehensive coverage of the parameter space detailed in Table 1, without demand for additional, computationally intensive PDE simulations. The enriched training dataset allowed the ML model to generalize more effectively, capturing subtle dependencies in nanoelectrode behavior. The combined approach (using PDE simulations and systematic data enhancement) implements a critical balance between computational efficiency and predictive accuracy, making it well-suited for the parameter estimation task. The supervised learning strategy underlying this workflow is sketched in Fig. 3 and detailed as follows:

- **Data segmentation:** The frequency-dependent behavior of capacitance is a critical factor in interpreting nanoelectrode array sensor measurements and predicting the capacitance response to analytes. At low frequencies, the system response is dominated by ionic migration and polarization effects, while at high frequencies, capacitive and electronic effects become more prominent. To ensure a realistic investigation, the hardware configuration directs our analysis toward using a spectrum between 2 MHz and 70 MHz, which is an optimal range for the chip (Widderhoven et al., 2018). In this work, the dataset is divided into

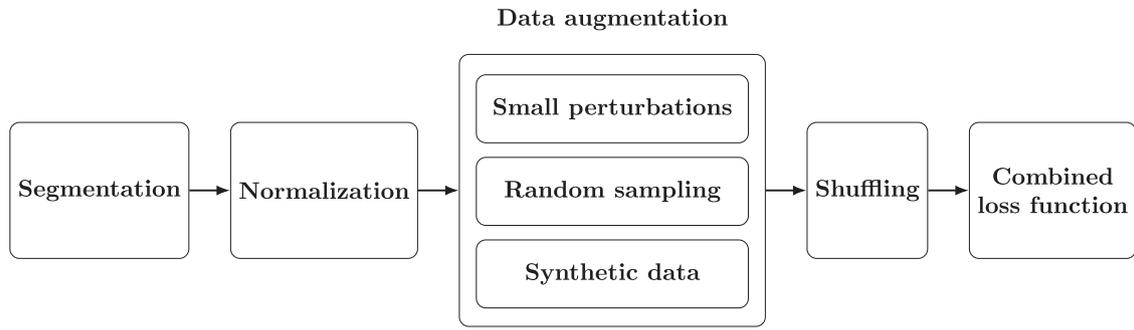


Fig. 3. Sketch of the supervised learning pipeline featuring structured data preprocessing and training steps. A combined loss function integrates least square error with custom weighting for optimal model training.

two subsets (low frequency, $f < 10$ MHz, and high frequency data, $f > 10$ MHz). The 10 MHz value is a natural mid-point and allows an effective partition between low-frequency values near the first electrolyte cut-off frequency and high-frequency values approaching the dielectric relaxation cut-off frequency of the electrolyte (Widdershoven et al., 2018). This segmentation allows the model to better capture frequency-specific patterns, improving its ability to generalize and predict the sensor behavior across the entire frequency spectrum.

- **Normalization:** A specialized normalization approach using the StandardScaler (Pedregosa et al., 2011,?) was applied separately for each frequency subset and output. The transformation is given by: $x_{\text{scaled}} = (x - \mu) / \sigma$, where μ is the mean and σ is the standard deviation of the feature, computed over the training dataset. The transformation ensures that the resulting feature values have a mean of 0 and a standard deviation of 1. This maintains the intrinsic distribution and variability of each feature, balancing the learning process and preventing any single feature from dominating due to scale differences. It enhances learning stability and convergence speed, which is crucial for precise performance in frequency-specific tasks.
- **Data augmentation:** The initial dataset of capacitance simulations by ENBIOS has been doubled in size by means of three distinct and equally weighted augmentation techniques (2850 capacitance maps each):
 - **Small perturbations** are introduced to simulate minor noise and inherent variability in sensor readings. The noise magnitude, σ_n , was set to 5% of the range of the feature, in turn defined as the difference between the feature's maximum and minimum values. By generating 2850 additional training samples with perturbed inputs and unchanged output labels, we enhance the robustness of the model to minor variability while preserving the physical consistency of the forward model. The outputs remain unchanged under the assumption that small input deviations do not significantly alter the capacitance response due to the smoothness of the underlying PDE model.
 - **Random sampling** was conducted within the feature ranges of the original dataset, broadening the input space and allowing the model to learn from a wider distribution of conditions. For each feature x , the new values were sampled uniformly between the minimum and maximum values of the feature in the original dataset. The corresponding labels were either interpolated or derived from the nearest original samples to maintain consistency with the data distribution.
 - **Synthetic data generation using GAN** was also utilized to expand the training set. The GAN framework consists of two components: a *generator*, which learns to produce realistic input samples that replicate the statistical distribution of the original dataset, and a *discriminator*, which

distinguishes between real and generated samples. Through iterative adversarial training, the generator improves its ability to create plausible input data that closely matches the original dataset. We adopted a lightweight feedforward GAN architecture, where both the generator and discriminator consisted of three fully connected layers with ReLU activations. The generator's input was a 3-dimensional latent vector sampled from a standard normal distribution. The generator's output was a 3-dimensional synthetic input sample representing (L, ϕ, ϵ_r) , bounded within the physical limits used in our study. The discriminator was trained to distinguish between real and synthetic input samples based on their statistical similarity to the original dataset.

The GAN was trained using the standard minimax adversarial loss, with the generator and discriminator updated alternately using the Adam optimizer. Convergence was monitored using both the discriminator accuracy and the Wasserstein distance between the real and synthetic input distributions. Once trained, the generator was used to produce new, realistic input samples. These synthetic samples were then passed through the pre-trained forward model (ML surrogate of the PNP-AC simulator) to obtain corresponding capacitance values, ensuring that the augmented dataset remained physically meaningful and consistent. The additional synthetic input samples generated by the GAN were paired with corresponding output labels obtained from a pre-trained physics-based model. This ensures that the generated data remains consistent with the underlying physical phenomena and retains its predictive accuracy. This method captures richer variations and improves the overall robustness of the ML model.

Among the data augmentation strategies used in this work, the introduction of small perturbations to the input features while keeping the corresponding output values unchanged requires careful justification. In our setting, the inputs — namely, analyte length, relative permittivity, and in-plane orientation angle — are physical parameters that govern the output capacitance through a smooth and continuous forward model. Small, controlled variations (on the order of 5% of the parameter range) in these inputs are not expected to produce abrupt or discontinuous changes in the corresponding output. The underlying physics modeled by the PNP-AC system exhibits stability and gradual sensitivity to such parameter changes, especially in regions where the forward model is locally linear or weakly nonlinear. Therefore, applying Gaussian perturbations to the inputs while keeping the original output labels allows the ML model to better generalize around each data point by encouraging local invariance. This approach also increases the effective size and diversity of the training set without requiring expensive new PDE simulations.

Although less common in high-dimensional vision tasks where labels are categorical and may shift under perturbation, this technique is well-motivated in physics-informed regression problems where smooth response surfaces allow for local label consistency. Moreover, we validated empirically that the introduced perturbations did not degrade model accuracy and instead improved generalization across unseen test cases. This aligns with recent findings in the field of scientific machine learning, where local consistency assumptions are increasingly used to balance data efficiency and model robustness (Wang and Perdikaris, 2021; Raissi et al., 2019).

This carefully designed augmentation strategy is pivotal in developing a resilient and adaptable ML model, capable of accurately predicting nanoelectrode capacitance across a wide range of operational scenarios. Indeed, small perturbations improve robustness to noise, random sampling expands the input space, and synthetic data captures intricate relationships within the system. We will show in Section 6 how the combined approach significantly enhances the model's ability to generalize to unseen conditions by using a more diverse and comprehensive dataset.

- **Data shuffling:** To mitigate sequential bias in the sensor capacitance data and promote robust learning, we apply data shuffling before each training epoch. Shuffling prevents the model from memorizing any specific sequence of capacitance values or augmentation types. This randomization ensures that the model learns meaningful, generalizable relationships across all data points rather than focusing on patterns associated with a particular data ordering. Furthermore, shuffling the augmented dataset prevents overfitting to specific patterns and ensures exposure to a diverse range of examples during training. To quantify the advantages of shuffling, we observed a reduction in the test set when shuffling was applied, demonstrating improved predictive accuracy. This improvement highlights the cost-effectiveness of shuffling, as it optimizes the learning process with minimal computational overhead while significantly boosting the model's robustness and reliability for realistic capacitance measurements.
- **Combined loss function:** To effectively train the model on sensor capacitance data, we employ a combined loss function that integrates the standard MSE with a custom, weighted component as follows:

$$\mathcal{L}_{\text{combined}} = \mathcal{L}_{\text{MSE}} + \lambda \sum_{i=1}^n w_i \cdot \mathcal{L}_i, \quad (7)$$

where \mathcal{L}_{MSE} is the standard mean squared error that evaluates the global prediction accuracy, \mathcal{L}_i is the individual loss associated with the specific output feature i , w_i is the custom weights assigned to each output feature based on its variability and λ denotes the scaling factor to balance the contributions of the MSE and the weighted components. The custom weights w_i are defined as:

$$w_i = \frac{\sigma_i}{\sum_{j=1}^n \sigma_j}, \quad (8)$$

where σ_i is the standard deviation (variability) of output feature i , and n is the total number of output features. This normalization prevents any single feature from dominating the training process. While data normalization standardizes both input features and output variables (i.e., capacitance) to zero mean and unit variance to stabilize model training, the custom weights w_i are used in the loss function to account for variability in the original (non-normalized) output features. All predictions are de-normalized (i.e., transformed back to real capacitance values) before being used in the Bayesian inversion framework.

This distinction resolves any apparent contradiction: the normalization step ensures balanced scaling of inputs, whereas the

combined loss function accounts for inherent output variability to emphasize features that are most critical for accurate capacitance prediction.

The combined loss function, when used alongside data augmentation and shuffling, refines the model's focus during training. Data augmentation increases diversity in the training set, shuffling ensures a randomized presentation of examples, and the weighted loss prioritizes critical output features. This integrated approach improves the model's generalization, balancing overall accuracy and feature-specific precision, thereby enhancing its robustness and reliability for predicting capacitance behavior across diverse operational conditions.

The fully connected DNN includes 4 input features (L , ϕ , ϵ_r , and f), 4 hidden layers with 128, 64, 32, and 16 neurons each, and 49 output labels (i.e., 7×7 pixel capacitance maps). Using the simulation outputs from ENBIOS, we implemented the ML workflow using *PyTorch* for efficient tensor operations and neural network construction, with the Adam optimizer employed for training. Data normalization was performed using *sklearn.preprocessing*, while *NumPy* was utilized for robust data augmentation. The training process was executed on a system equipped with an Intel Core i7-13700H CPU and 32 GB of memory, ensuring an efficient and streamlined computation pipeline.

Table 2 presents the CPU time required for the training and testing phases, comparing scenarios with and without data augmentation. A dramatic reduction of the computational burden with respect to physics-based numerical simulations is achieved (compare Table 2 to Fig. 2). A reduction in the time for training and testing is observed for increasing frequencies due to the physical system's tendency to exhibit smaller gradients and weaker non-linearities. This behavior is likely due to the reduction in Debye screening effects, which entails a smaller capacitance sensitivity to the analyte's exact position, leading to simpler patterns in the dataset, and eventually requiring fewer computational resources for training and testing because the ML framework processes and predicts less intricate patterns.

To evaluate the trade-off between computational efficiency and model accuracy, we define three training scenarios: using 100%, 75%, and 50% of the original dataset, each tested with and without data augmentation. For the reduced-data cases, 75% or 50% of the full dataset was randomly selected to maintain coverage of the parameter space. We then applied the same augmentation strategy to these subsets. This setup (see Section 6) allows us to assess how well the model performs with fewer physical simulations, highlighting the balance between computational cost and predictive accuracy. It also demonstrates the value of data augmentation in maintaining performance when training data is limited.

Once trained, the ML model predicts sensor behavior with dramatically reduced computational costs compared to the FEM numerical simulations. This efficiency enables seamless integration with the Bayesian inversion framework, allowing for the rapid and accurate identification of unknown parameters within the system. This aspect is described in the following section.

5. The Bayesian inversion parameter estimation framework

Bayesian inversion is a robust statistical approach for extracting the physical and geometrical parameters of the analytes by integrating prior knowledge with observational data through Bayes' theorem, especially well-suited for nonlinear and high-dimensional models where traditional optimization techniques may falter.

A common implementation of Bayesian inversion employs Markov Chain Monte Carlo (MCMC) methods to approximate the posterior distribution of model parameters. These methods excel in handling non-Gaussian and nonlinear problems, providing flexibility in exploring complex parameter spaces (Cossettini et al., 2019; Stadlbauer et al., 2019).

Table 2

CPU time for the ML algorithm. The last column denotes the cumulative CPU time across ten different frequencies (2–70 MHz). The time unit is seconds..

Frequency [MHz]	2	6.54	31.77	70	2–70
Execution time for 100% data (with augmentation)	53	49	32	25	315
Execution time for 100% data (w/out augmentation)	32	29	19	12	196
Execution time for 75% data (with augmentation)	38	35	22	19	225
Execution time for 75% data (w/out augmentation)	24	22	15	9	145
Execution time for 50% data (with augmentation)	26	23	17	13	160
Execution time for 50% data (w/out augmentation)	17	16	10	6	110

5.1. Probabilistic model definition

We consider the following probabilistic framework:

$$\mathcal{M} = \mathcal{Q}(x, z) + \epsilon, \quad (9)$$

where $\mathcal{Q}(x, z)$ represents the computational model (here the AC-PNP equations) parameterized by z , which belongs to the random field \mathcal{Z} . The observational error ϵ is assumed to be Gaussian independent and identically distributed (i.i.d.), i.e., $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, where \mathcal{N} denotes the normal distribution, σ^2 is the error variance. The measured data $\mathcal{M} = \text{obs}$ lead us to define the conditional density as:

$$\pi(\text{obs}) = \int_{\mathbb{R}^n} \pi(\text{obs}|z)\pi_0(z)dz. \quad (10)$$

The central outcome of Bayesian inversion is the posterior probability density $\pi(\text{obs})$, which represents updated beliefs about the parameters after accounting for the data. This posterior density incorporates both the prior distribution $\pi_0(z)$ and the likelihood $\pi(m|z)$, ensuring a balance between prior information and new observational evidence. The posterior density for the parameter z , given an observation m , is computed as:

$$\pi(z|m) = \frac{\pi(m|z)\pi_0(z)}{\pi(m)}, \quad (11)$$

where $\pi_0(z)$ is the prior density. Using the probabilistic model above, the likelihood function is defined as:

$$\pi(\mathcal{M}|z) = L(z, \sigma^2|\mathcal{M}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\mathcal{E}\mathcal{E}_{\mathcal{M}}}{2\sigma^2}\right), \quad (12)$$

where the misfit between the model predictions and the observed data, $\mathcal{E}\mathcal{E}_{\mathcal{M}}$ is:

$$\mathcal{E}\mathcal{E}_{\mathcal{M}} = \sum_{j=1}^n [\mathcal{M}_j - \mathcal{Q}_j(x, z)]^2, \quad (13)$$

and is minimized during the parameter estimation process.

5.2. MCMC methods and DRAM

For estimating the posterior distribution, different MCMC methods can be considered (Noii et al., 2022). Among these, the *Delayed Rejection Adaptive Metropolis* (DRAM) algorithm is particularly notable (Haario et al., 2006). DRAM builds on the Metropolis–Hastings method (Hastings, 1970) by incorporating two key features: delayed rejection and adaptive updates. Delayed rejection allows re-evaluating rejected proposals, increasing acceptance in low-probability regions (Green and Mira, 2001). Adaptive updates adjust the proposal distribution based on the history of accepted moves, improving convergence and efficiency (Haario et al., 1999). This combination makes DRAM well-suited for high-dimensional, complex posterior distributions. For a detailed implementation of the MCMC algorithm in sensor parameter estimation, we direct interested readers to Khodadadian et al. (2020b) and Stadlbauer et al. (2019).

1. Adaptive Metropolis: DRAM dynamically adjusts the covariance matrix of the proposal distribution based on the sample history. This

adaptation ensures efficient sampling and helps the algorithm handle high-dimensional or correlated parameter spaces. The updated covariance matrix at iteration k is:

$$\Sigma_k = \frac{1}{k-1} \sum_{i=1}^{k-1} (z_i - \bar{z}_k)(z_i - \bar{z}_k)^T + \epsilon I, \quad ,$$

where \bar{z}_k is the running mean of the samples, and ϵ is a small positive constant added for numerical stability. To generate a new proposal z_k from the previous candidate z_{k-1} , a multivariate normal distribution is used:

$$z_k \sim \mathcal{N}(z_{k-1}, \Sigma_k).$$

Using the Cholesky decomposition of Σ_k ($\Sigma_k = LL^T$), the proposal can be expressed as:

$$z_k = z_{k-1} + Lv,$$

where $v \sim \mathcal{N}(0, I)$ is a standard normal random vector.

2. Delayed Rejection: If an initial proposal z_k is rejected, DRAM generates a secondary proposal z'_k with a modified covariance structure, improving the exploration of the parameter space. The secondary proposal z'_k is drawn from a distribution $q_2(z'_k|z_{k-1})$, which can adjust both the covariance and the mean based on the rejected state z_k . The computation of z'_k involves:

- Adjusting the covariance matrix to scale down the step size:

$$\Sigma'_k = \rho \Sigma_k,$$

where $\rho \in (0, 1)$ is a scaling factor.

- Generating the proposal:

$$z'_k = z_{k-1} + L'v',$$

where L' is the Cholesky decomposition of Σ'_k ($\Sigma'_k = L'L'^T$), and $v' \sim \mathcal{N}(0, I)$ is a standard normal random vector.

- Optionally, incorporating the rejected state z_k to improve exploration:

$$z'_k \sim \mathcal{N}(z_k + \delta(z_k - z_{k-1}), \Sigma'_k),$$

where δ is a small step-size parameter that biases z'_k slightly away from z_k .

The acceptance probability for the second stage is:

$$\alpha_2(z_{k-1}, z'_k) = \min\left(1, \frac{\pi(z'_k)q_1(z'_k|z_k)q_2(z_{k-1}|z'_k)(1 - \alpha_1(z'_k, z_k))}{\pi(z_{k-1})q_1(z_k|z_{k-1})q_2(z'_k|z_{k-1})(1 - \alpha_1(z_k, z_{k-1}))}\right), \quad (14)$$

where q_1 and q_2 are the proposal densities for the first and second stages, respectively, and α_1 is the acceptance probability for the first stage. The algorithm is shown in Fig. 4 where the developed ML framework is used to compute the capacitance in the sensing points.

6. Results and discussion

In this section, we first evaluate the accuracy of the ML model by analyzing its MSE error and validating its performance against benchmark results. Once the accuracy is established, we integrate the model

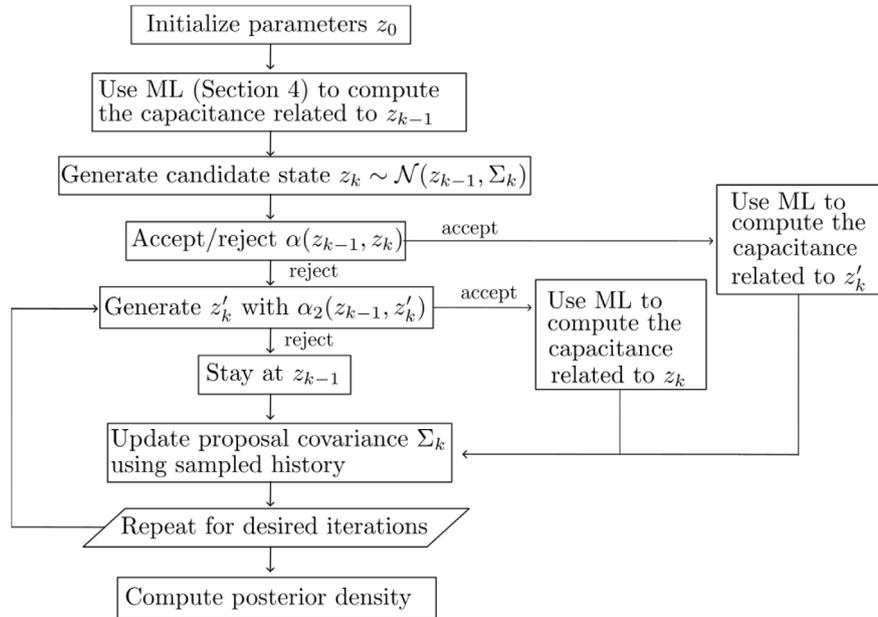


Fig. 4. Flowchart of the DRAM algorithm using the ML model to compute capacitance.

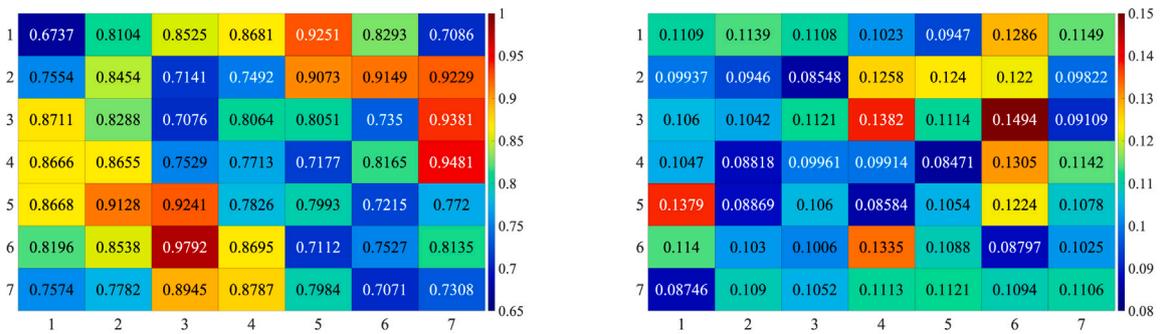


Fig. 5. Heatmaps of the relative error for ML model predictions on sensor capacitance data over a 7×7 test array for the initial model's without data augmentation (left, mean = 0.817 ± 0.0766 at one sigma), and for the improved model featuring targeted data augmentation (right, mean = 0.108 ± 0.0147 at one sigma). The training and testing datasets span the whole 2–70 MHz frequency range.

into a Bayesian inversion framework to efficiently identify the unknown parameters of the analyte. We also briefly discuss the computational and training aspects of the ML approach, emphasizing its practical advantages.

6.1. Machine learning model for the sensor response

The dataset is divided into three subsets: training (80%), validation (10%), and testing (10%). To study the ML accuracy, we compute the relative capacitance error, calculated as the difference between the predicted capacitance values from the ML model and the actual capacitance values, normalized to the actual values, for 50 samples randomly chosen out of the 10% saved for testing.

Then, for each electrode location in the 7×7 electrode maps, we average over samples and frequencies, thus generating a 7×7 error map and a total of 49 error values, representing the performance of the ML model over frequencies and samples in a spatially resolved manner. By restricting the frequency range, we also obtain error maps representing the ML model performance in specific frequency ranges, as discussed in Section 6.2.

Fig. 5 displays the relative error distribution for the ML model's performance during testing. Each cell represents the relative error with a color intensity reflecting its magnitude. In the left panel (the one for the model trained without data augmentation, i.e. without the

center block in Fig. 3), the relative error distribution shows notable variation and high error values. This indicates areas where the model's predictions without data augmentation are less accurate, likely due to local complexity in the data or insufficient representation during training. These discrepancies may stem from high variability in the input features, challenging parameter combinations, or the model's limited ability to generalize. The right panel reports the results for the complete model in Fig. 3, and demonstrates error maps free of spatial patterns (as expected), and a $7 \times$ reduction in error and standard deviation across the grid. This indicates improved predictive accuracy and consistent performance across the array. The adopted strategies collectively enhance the model's ability to generalize, reduce the variability in its predictions, and effectively mitigate the impact of the analyte's rotations and other sources of variability, thus achieving a more uniform and accurate prediction across all test configurations.

Fig. 6 presents heatmaps of the relative error between ML model predictions and reference sensor capacitance values, evaluated across a 7×7 test array for two reduced dataset, i.e., 75%, and 50%. In each case, data augmentation was applied to enhance model generalization. The average relative error increases as the size of the original training dataset decreases: 0.108 for 100% (see Fig. 5), 0.1229 for 75%, and 0.1773 for 50%. These results are consistent with expectations, as reducing the training data limits the model's exposure to variability in the physical parameter space (e.g., length, permittivity,

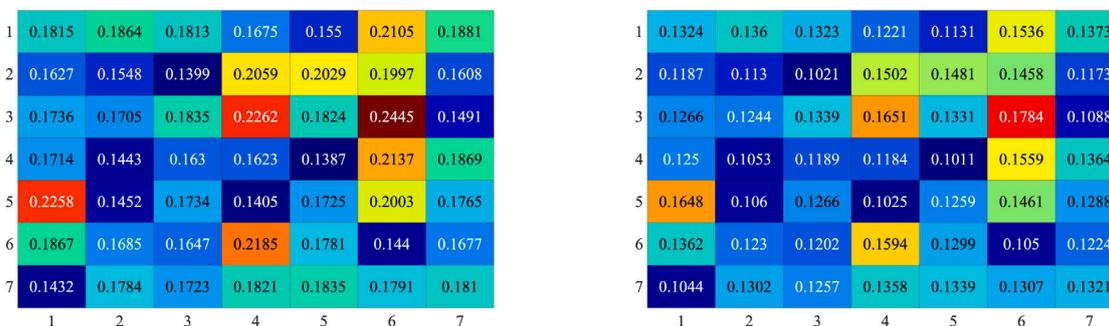


Fig. 6. Heatmaps of the relative error for ML model predictions on sensor capacitance data over a 7×7 test array for the initial model's using 50% of the data (left, mean = 0.1773 ± 0.009 at one sigma), and for the 75% of the data (right, mean = 0.1229 ± 0.0087 at one sigma). The training and testing datasets span the whole 2–70 MHz frequency range.

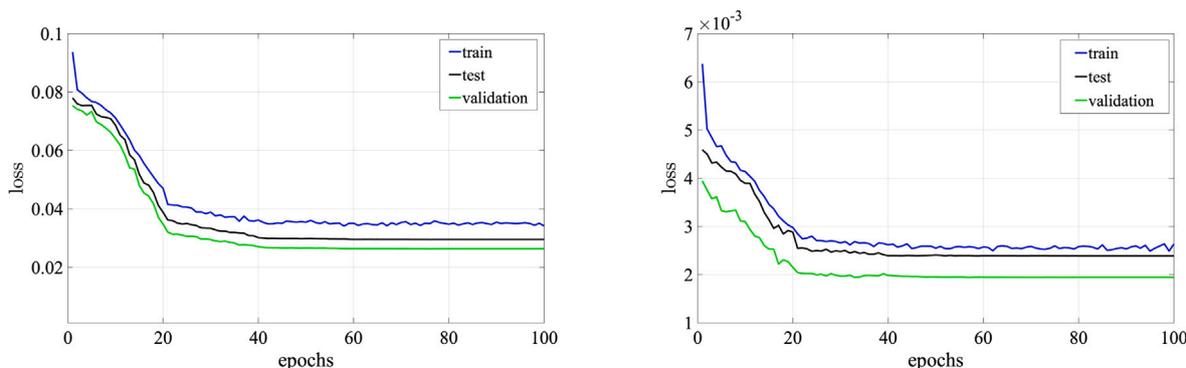


Fig. 7. Evolution of the loss functions for training, testing, and validation for ML models without (left panel) and with (right panel) data augmentation. The training, testing, and validation datasets span the whole 2–70 MHz frequency range.

and orientation). While data augmentation helps to enrich the training distribution, it cannot fully replace the diversity and coverage of the original simulations. Nevertheless, the relatively modest increase in error when using 75% of the data suggests that the model retains strong predictive capability even under constrained data conditions. The more pronounced error in the 50% case reflects a reduced ability to capture finer patterns in capacitance variations, though the model still performs reasonably well given the significant reduction in simulation cost.

Table 2 and Fig. 6 jointly illustrate the trade-off between computational efficiency and model accuracy. Training with 100% of the data achieves the lowest average relative error (0.108) but requires the highest CPU time (315 s with augmentation). Reducing the training data to 75% and 50% lowers the training time to 210 s and 160 s, respectively, while increasing the relative error moderately to 0.1229 and 0.1773. These results demonstrate that significant computational savings — up to 49% — can be achieved with only a marginal drop in accuracy, particularly in the 75% case. This scenario represents a promising trade-off for rapid prototyping, model iteration, or deployment in resource-constrained environments. However, as the remaining parts of this study aim to maximize prediction accuracy and ensure the highest fidelity in parameter estimation, we continue with the full 100% dataset in all subsequent analyses.

Fig. 7 compares the loss function evolution for the training, testing, and validation phases, for ML models without (left) and with (right) data augmentation. Without data augmentation, the loss functions show a slower initial decay and converge to higher final loss values, confirming limited generalization and accuracy. Conversely, with the proposed improvements the loss function decreases more rapidly, and the final loss values are significantly lower across all data sets. This improvement highlights the effectiveness of data augmentation. By introducing augmented data, the ML model is exposed to a more diverse set of training examples, enabling it to learn broader patterns and

mitigate overfitting. Furthermore, the close alignment of the validation and test loss curves with the training loss curve in the augmented case demonstrates that the model is well-generalized and avoids underfitting or overfitting.

6.2. Frequency effect

The nanoelectrodes' ΔC generated by the analytes are quite sensitive to the signal frequency. Fig. 8(a) illustrates this aspect showing typical spectra of $-\Delta C$ (absolute value, left, and normalized value, right) for the two first neighbors of electrode (0,0) (the one having the largest $|\Delta C|$) on the row ((0,+1)) and column ((+1,0)) directions for an elongated analyte featuring $\phi = 0^\circ$, $L = 1000$ nm, i.e barely overlapping the neighbor electrode along x but not the one along y .

At the chosen electrolyte salinity, the 2–70 MHz spectrum covers a large portion of the frequency range between the lowest cut-off frequency of the sensor response (which is largely dependent on the electrolyte resistance; ≈ 1.7 MHz in our case), and the dielectric relaxation cut-off frequency of the electrolyte ($\sigma_{el}/\epsilon_{el} \approx 300$ MHz in our case), beyond which Debye screening is largely suppressed. Consequently, decreasing $\Delta C(f)$ values are observed. In the ≈ 2 –10 MHz range, the capacitance is dominated by the large electrical double-layer component; its change reflects perturbations near the surface where polarization effects are slow but remarkable and sensitive to the proximity between electrodes and the analyte. In the 10–70 MHz range, instead, Debye screening weakens and the capacitance keeps decreasing due to the penetration of the AC electric field in the bulk of the electrolyte. Configurational changes have a reduced impact on the capacitance response and the sensitivity of the capacitance to perturbations next to the nanoelectrode relative to the one far from it decreases appreciably. In other words, the response is less sensitive to the analyte configuration and more to its volume at high frequency

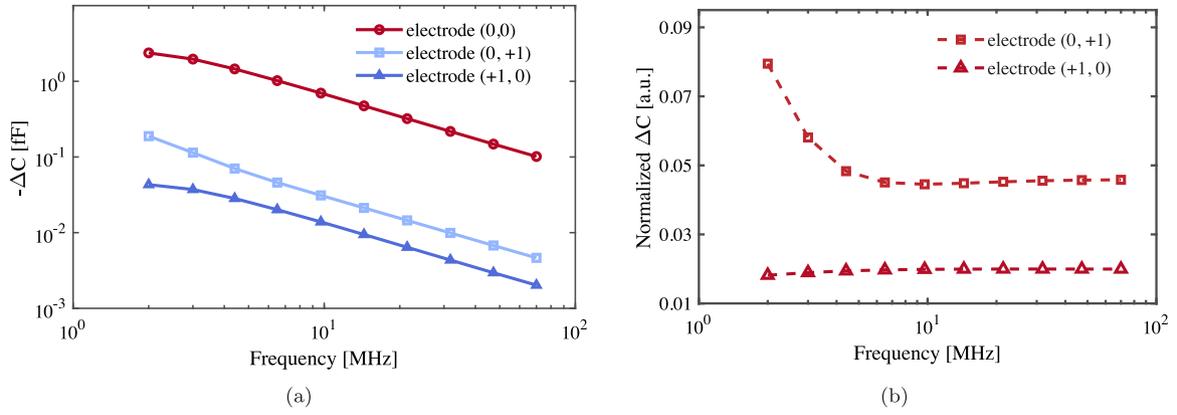


Fig. 8. (a) Representative example of the spectral response of three electrodes below an analyte ($L = 1000$ nm, $\phi = 0^\circ$, $\epsilon_r = 1$. All the other parameters as in Table 1). Electrode (0,+1) is the first neighbor to the central electrode (0,0) in the same row, thus located at 600 nm distance, while electrode (+1,0) is on the same column, thus at a distance of 890 nm. See also Fig. 1(b). (b) $\Delta C_{0,+1}$ and $\Delta C_{+1,0}$ normalized to $\Delta C_{0,0}$.

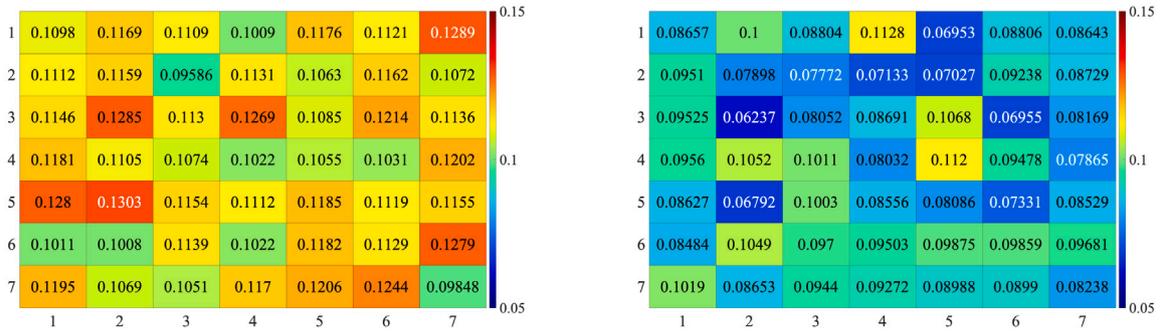


Fig. 9. Heatmaps of relative capacitance error in ML model predictions for a 7×7 nanoelectrodes array when the model is trained on the full 2–70 MHz spectrum and tested only on the low-frequency data (left, $f = 2$ –10 MHz, mean = 0.1134 ± 0.0086 at 1 sigma), and the high-frequency data (right, $f = 10$ –70 MHz, mean = 0.0887 ± 0.0117 at 1 sigma).

than at low frequency. As a result, the frequency dependence of the ΔC is more diverse among electrodes at low frequency, and more similar at high frequency. This is visible in the left plot of Fig. 8(b), and even more evident in the normalized plot on the right. Therefore, the effect of frequency on capacitance is a key consideration in the interpretation of nanoelectrode array sensor measurements and in predicting the capacitance response to analytes.

To investigate this aspect further, Fig. 9 presents the heatmaps of the relative error in ML model predictions for sensor capacitance averaged over five frequencies in the 2–10 MHz range (left) and five in the 10–70 MHz range (right). The dataset encompassing the whole 2–70 MHz spectrum was used for training. Both maps are still free of geometric patterns, as they should be with a well-balanced model, and the relative error values are comparable to those in the right panel of Fig. 5. In the 2–10 MHz frequency range, the model's relative errors are slightly larger than at high frequency (≈ 0.1 –0.15 as opposed to ≈ 0.05 –0.1). This difference may be attributed to the model's sensitivity to the complex capacitance behavior at low frequencies, as illustrated in Fig. 8(a). The smaller error in the right panel suggests that capacitances are more predictable at higher frequencies, as indeed observed in Fig. 8(b). This is due to a more uniform sensitivity to fluctuations and disturbances occurring at short and large distances from the electrodes. Consequently, the model encounters less complexity and more effectively generalizes across different configurations.

Fig. 10 compares the loss functions for training, testing, and validation when the model is trained on the full frequency spectrum (2–70 MHz) and separately tested on low-frequency data (left, 2–10 MHz) and high-frequency data (right, 10–70 MHz); that is, the same conditions as in Fig. 9. In all cases the losses decrease rapidly and reach full convergence within the explored 100 epochs range; in practice within

a few tens of epochs. The asymptotic values are slightly smaller in the high-frequency testing case (right plot), consistent with the previous discussion about Fig. 10.

Additionally, we investigate the performance of the ML model trained on capacitance data and tested on errors constrained to either the low- ($f < 10$ MHz) or the high-frequency range ($f > 10$ MHz), rather than the entire range (2–70 MHz) as it was in the previous case. This approach assumes that capacitance data are available only in either one of the ranges, but not both. This constraint is relevant in practice, as measuring a wide band ΔC spectrum is not trivial nor common (see overview of platforms in Section 2), and may face technological limitations.

Fig. 11 demonstrates the relative error distributions for scenarios where the ML model is trained and tested exclusively on either low-frequency (left) or high-frequency (right) capacitance data, respectively. Both heatmaps highlight again the absence of any clear spatial pattern. The average errors over the array are slightly larger than in the previous case, where the entire frequency range (2–70 MHz) was used for training, Fig. 9. Therefore, using the full frequency range for training improves model reliability and allows the model to capture underlying trends and nonlinearities more effectively. The degradation of the ML model errors when restricting the frequency range for training and testing is not critical in the chosen range and is still quite acceptable. However, having results for both low and high frequencies is key since it ensures sensitivity to bulk and surface phenomena, enhances robustness against noise and uncertainty, and improves sensitivity. This dual-range approach provides better insights, enhances physical understanding but also strengthens the reliability of the ML model predictions.

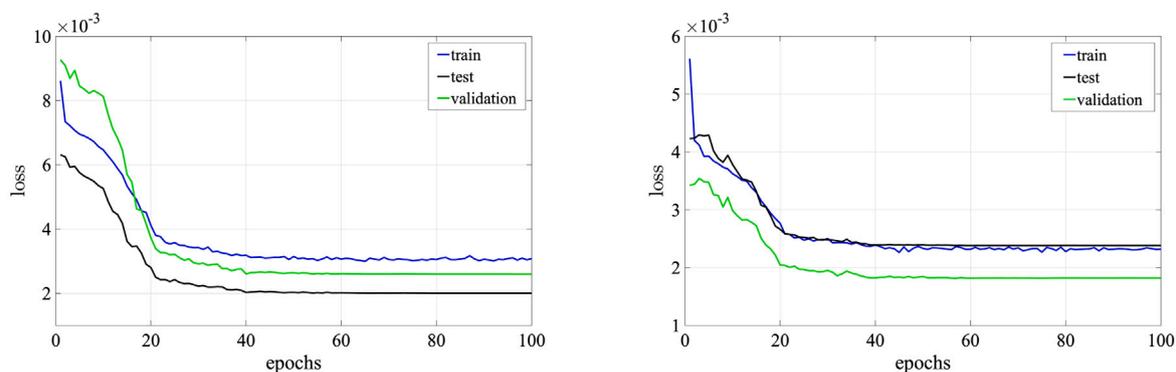


Fig. 10. Loss functions for training, testing, and validation when the model is trained with data on the full 2–70 MHz frequency spectrum and tested on low-frequency data (left, 2–10 MHz) and high-frequency data (right, 10–70 MHz) only; i.e. the same conditions as in Fig. 9.

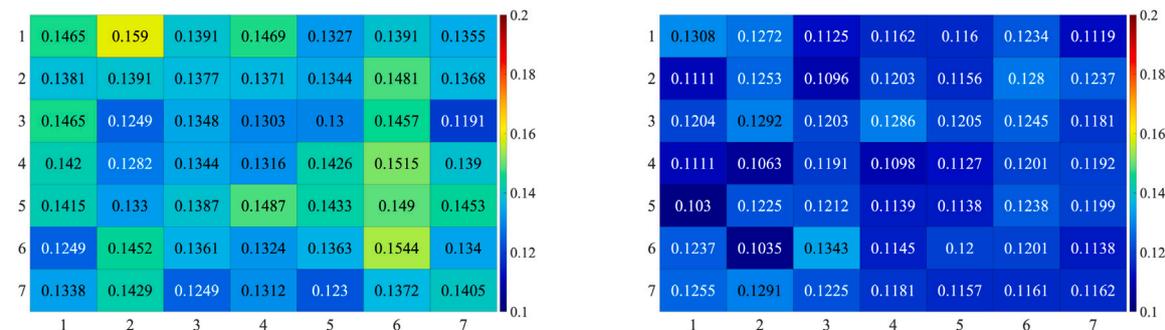


Fig. 11. Heatmaps of relative capacitance error in ML model predictions for 7×7 nanoelectrodes images when the model is trained and tested on the low-frequency data (left, $f = 2\text{--}10$ MHz, mean = 0.1381 ± 0.0083 at 1 sigma), and on the high-frequency data (right, $f = 10\text{--}70$ MHz, mean = 0.1188 ± 0.0069 at 1 sigma).

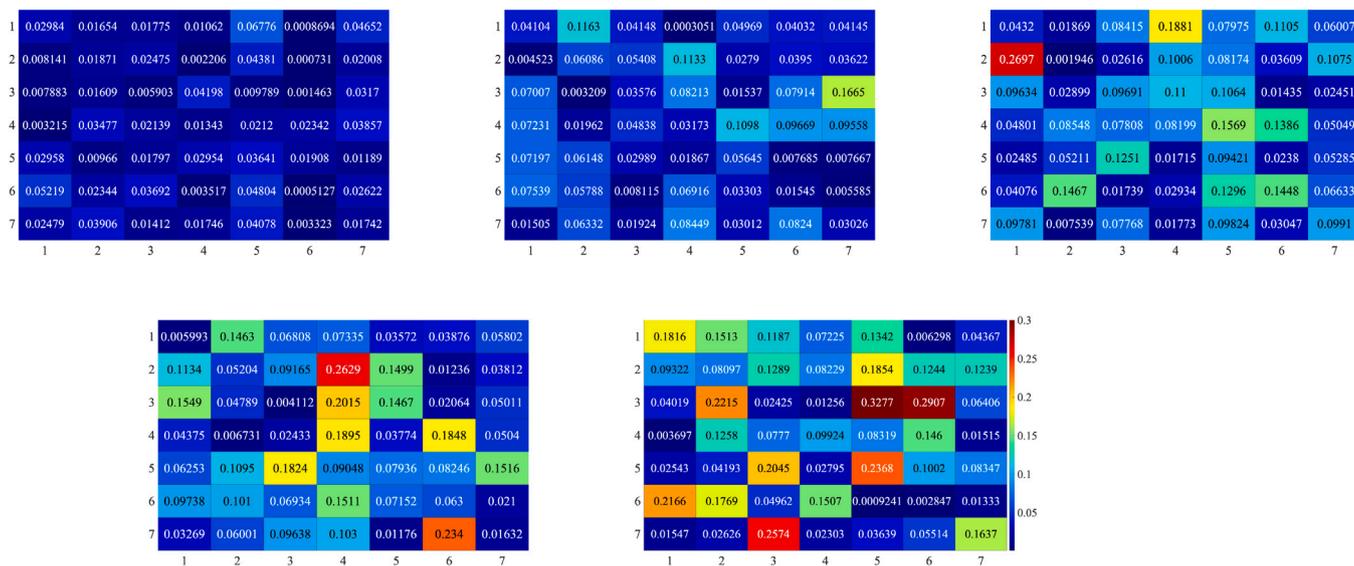


Fig. 12. Heatmaps of the relative capacitance error in the ML model's extrapolation to higher frequencies. The model was trained using data from 2 to 70 MHz and applied to estimate capacitance at (top left) $f = 80$ MHz (mean = 0.0221 ± 0.0156), (top middle) $f = 90$ MHz (mean = 0.0497 ± 0.0355), (top right) $f = 100$ MHz (mean = 0.0759 ± 0.0530), (bottom left) $f = 125$ MHz (mean = 0.0856 ± 0.0635), and (bottom right) $f = 150$ MHz (mean = 0.1014 ± 0.0821). The extrapolation was performed for a fixed analyte length $L = 600$ nm, orientation $\phi = 90^\circ$, and permittivity $\epsilon = 3$. The color bar (bottom right) applies to all heatmaps.

As a last step, we examine the ability of the ML model to generalize the prediction of the ΔC maps beyond the frequency range available for training. This analysis is particularly interesting since increasing the measuring frequency beyond 100 MHz is challenging (Cossettini et al., 2020), as also highlighted by the state-of-the-art review reported in Section 2, but beneficial for a deeper probing of the electrolyte, especially with high salt concentrations. To this end, Fig. 12 presents

heatmaps illustrating the relative error in capacitance estimation when the ML model, trained on frequencies between 2 and 70 MHz, is used to extrapolate capacitance at higher frequencies (80, 90, 100, 125, and 150 MHz). Each heatmap corresponds to a different extrapolated frequency and quantifies the error distribution across spatial coordinates. As the extrapolation frequency increases, the mean relative error also rises, suggesting a decline in model accuracy at higher frequencies.

Nonetheless, the average error over the array remains $\leq 10\%$ even when the excitation frequency is as large as $2\times$ the maximum frequency in the training set. This trend highlights the challenge of extending ML-based predictions beyond the training range, particularly in the context of nanoscale capacitance spectroscopy.

6.3. Nanoelectrode parameter estimation

The three unknown parameters (ϵ_r , L , ϕ) play distinct roles in determining the capacitance and are critical for accurately characterizing the analytes and the metrological performance of the system in biological detection and environmental monitoring.

- 1. Permittivity:** This parameter is directly linked to the material properties of the analyte and governs how the electric field interacts with it. Measuring permittivity at nanoscale dimensions is very challenging due to the complex interactions between the probe, the surrounding media, and the analyte.
- 2. Analyte length:** The length of the cylindrical analyte impacts the distribution of the electric field across the electrodes and, consequently, the capacitance maps. Analyte length is difficult to measure precisely in real time since high-resolution physical measurements inside the microfluidic chamber are essentially impossible at the nanoscale explored here.
- 3. Analyte in-plane angle:** The rotation angle of the cylindrical analyte relative to the nanoelectrodes is critical for defining the spatial alignment of the system, which directly affects the capacitance and the measured signal. Accurately determining this angle is particularly challenging when the analyte dimensions are comparable to the electrode pitch, as in our case.

Additional challenges stem from:

- the experimental noise, particularly at nanoscale dimensions where electronic and thermal fluctuations can dominate.
- The non-linear relationship between the unknown parameters and the observed capacitance, due to the complex AC electric field distribution around the analyte.
- The parameter space may exhibit multiple plausible solutions, requiring methods to explore complex, multimodal distributions.

By applying Bayesian inversion, we ensure that all available information — both new data and prior knowledge — is utilized, leading to more reliable and interpretable parameter estimates for the nanoelectrode system. The forward model required for Bayesian inversion was replaced by the trained ML framework validated in the previous sections.

To generate a representative and informative dataset for training the ML model, we employed a structured sampling strategy over the parameter space defined by the analyte length, relative permittivity, and in-plane rotation angle. The sampling was designed to uniformly cover the physically relevant ranges of these parameters. We used a combination of grid-based sampling and random perturbations to ensure both dense coverage and diversity within the parameter space. Additionally, to improve the generalization capability of the model and reduce overfitting, we applied data augmentation techniques including additive Gaussian noise, uniform resampling, and synthetic data generation via GANs. This hybrid approach resulted in a dataset that captures both the global variability and local subtleties of the input-output relationships, thereby enhancing the robustness and predictive accuracy of the supervised learning framework. The final dataset supports efficient training and reliable parameter estimation across a wide range of physical scenarios. At present, the model is trained on fixed-resolution simulations. Possible advantages in adopting resolution-independent operator learning (Jiang et al., 2024) and discretization-agnostic Bayesian inference (Selig, 2014) are discussed in Section 7.

Table 3

Features of the four test cases considered for testing the accuracy of the integrated ML model and Bayesian inversion procedure. The parameter ranges are given in columns 2-3; the true values in columns 4-7.

Parameter	Min	Max	Case 1	Case 2	Case 3	Case 4
ϵ_r	1	6	3.5	2	2.5	4
L [nm]	100	1000	700	300	500	900
ϕ [°]	0	165	70	80	40	110

Table 4

Summary of posterior statistics for each case and parameter (mean, standard deviation (Std), 95% confidence interval (CI)). The true values are shown in Table 3. The reported results are computed after burn-in removal.

Case	Parameter	Mean	Std	95% CI
Case 1	L [nm]	700.46	4.46	[700.10, 700.81]
	ϵ_r	3.55	0.03	[3.54, 3.55]
	ϕ [°]	69.18	2.06	[69.04, 69.32]
Case 2	L [nm]	313	4.75	[312.6, 313.5]
	ϵ_r	2.02	0.021	[2.024, 2.027]
	ϕ [°]	78.26	2.38	[78.07, 78.46]
Case 3	L [nm]	493.6	9.28	[492.88, 494.33]
	ϵ_r	1.95	0.04	[1.95, 1.96]
	ϕ [°]	38.90	1.22	[38.81, 38.99]
Case 4	L [nm]	889.37	5.49	[888.87, 889.88]
	ϵ_r	3.93	0.04	[3.92, 3.94]
	ϕ [°]	110.30	2.78	[110.09, 110.51]

6.4. Bayesian inversion performance

To observe the accuracy of the Bayesian inversion and the developed ML framework, we consider four representative cases with different combinations of analyte parameters. By isolating and combining parameter variations in these cases, the framework assesses how well the trained model handles variability and recovers true parameter values across wide ranges of input features. The chosen test cases (see Table 3) balance complexity with computational feasibility, ensuring that the initial evaluation remains manageable. A uniform prior density was employed with minimum and maximum values as detailed in Table 3, corresponding to the complete absence of prior knowledge on the analyte parameters.

Fig. 13 presents the results of applying the integrated ML and Bayesian inversion framework to estimate the analyte parameters in Cases 1–4. Convergence of the chains toward the true values given in Table 3 (top row of the Figure) signifies that the algorithm successfully identifies the parameters. The subsequent rows display the posterior probability distributions for each parameter, providing insights into the accuracy and confidence of the estimates.

To provide a comprehensive analysis of the Bayesian inversion results, Table 4 summarizes the key statistical measures for each parameter (L , ϵ_r , and ϕ) across all four test cases. For each parameter, we report the posterior mean, standard deviation, and the 95% confidence interval, computed from the MCMC samples. These intervals represent the range within which the true parameter values are expected to lie with 95% probability. This quantitative evaluation offers a clear view of the reliability and uncertainty associated with the inferred parameters. The results confirm that, despite inherent noise and model approximations, the proposed ML and Bayesian framework achieves consistent and accurate parameter estimation across different physical configurations.

Across all four cases, the results show that the median of the posterior density consistently aligns closely with the reference values for all parameters. This indicates that the ML+Bayesian inversion framework can identify parameter estimates that are centered within less than 10% of the true values summarized in Table 3. The post-burn-in, posterior distributions (azure) and their Gaussian fits (black solid line) provide evidence of the uncertainty in these estimates; a piece of very

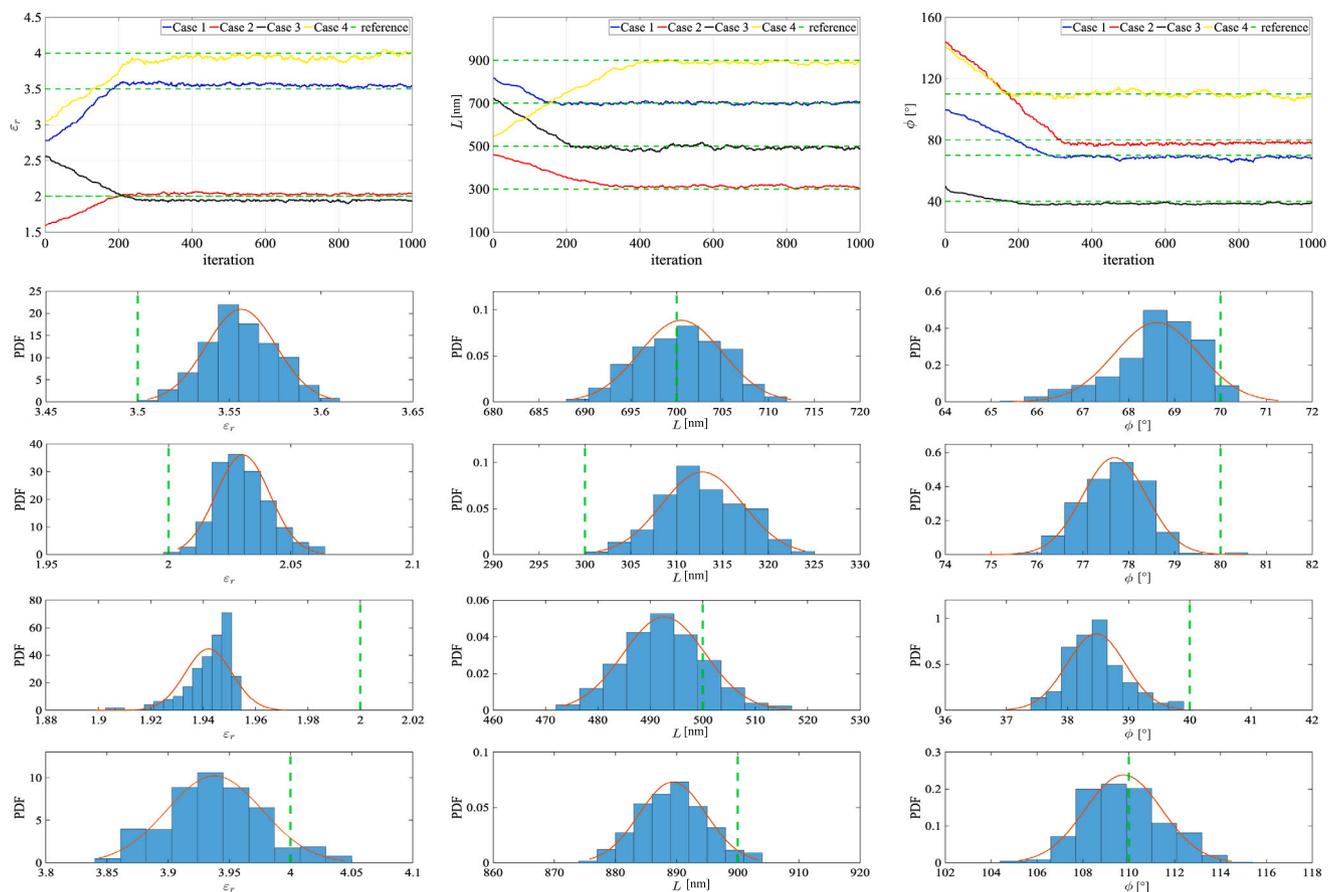


Fig. 13. MCMC trace plots (top row) and post-burn-in posterior distributions (rows 2–5) for the estimated parameters: relative permittivity ϵ_r (first column), analyte length L [nm] (second column), and orientation angle ϕ [°] (third column), across four test cases (Case 1 to Case 4, from second to fifth row). The green dashed lines indicate the truth values (see Table 3). Posterior distributions reflect the uncertainty and variability in the inferred parameters after the burn-in period, with corresponding fitted normal distributions overlaid in red.

relevant information for practical applications. While the medians are accurate, the presence of broader or skewed distributions is visible; for instance, those of the analyte length, L , in Case 3 and Case 4, reflecting increased uncertainty in parameter estimation. This is likely due to challenges such as reduced sensitivity of capacitance to certain parameters (e.g., ϕ) or the limited representation of these parameter combinations in the training data. Despite these issues, the closeness of the medians to the references demonstrates that the framework can reliably estimate central tendencies, even with small analytes featuring dimensions comparable to the nanoelectrode pitch.

As a final consideration, it is worth noting that, in the present framework, the ML model is employed as a deterministic surrogate for the forward simulation, enabling fast and accurate predictions of sensor behavior. However, this setup does not explicitly account for the predictive uncertainty of the ML model in the Bayesian inversion process. As a result, the posterior distributions presented here reflect uncertainties from the measurement noise and the prior, but not from the surrogate model itself.

7. Conclusions

This work investigated a viable strategy to empower innovative nanoelectrode array sensors with detection and metrological characterization algorithms for elongated analytes modeled as cylinders across a broad range of conditions and close to the spatial resolution limits of the sensor.

A combined ML and DRAM-MCMC Bayesian inversion framework is developed to this end. In particular, by leveraging carefully tailored

data augmentation techniques, the trained ML model provided a fast and accurate surrogate for the forward model, enabling efficient sampling of posterior distributions while significantly reducing reliance on costly, physics-based AC-PNP computations. The relative error in the capacitance predictions remains low even when segmentation of the frequency spectrum or extrapolation outside the frequency band of the training samples is necessary. The DRAM algorithm ensured robust convergence to true parameter values. This workflow reduced computational overhead while preserving accuracy, providing a practical solution for high-dimensional, nonlinear problems like the AC-PNP model.

The integrated ML + Bayesian inversion framework demonstrates strong performance in parameter estimation, with MCMC chains converging to reference values and posterior distributions aligning closely with true parameters in most cases. Despite challenges in specific cases with high variability or dependencies, the algorithm proved to be a computationally efficient and accurate tool for solving parameter estimation problems over multidimensional and broad intervals.

These results validate the framework's success in combining ML with Bayesian methods to replace computationally expensive AC-PNP (and in more general terms, PDE) simulations effectively. While the current version of the model has been trained for elongated bio-analytes in physiological electrolytes, and applications to less salty electrolytes (e.g., fresh waters or plant fluids) would require retraining, the strategy adopted and the combination of models and algorithms remain valid. Future improvements, such as incorporating more data or refining the ML model, could address these limitations.

Furthermore, while the DRAM algorithm provides robust posterior sampling, its computational cost remains a limiting factor. The use of

diffusion-based models to accelerate MCMC convergence by learning transition dynamics (Hunt-Smith et al., 2024), and introduce symmetry-aware sampling for efficient Bayesian neural inference (Wiese et al., 2023), could improve scalability and sampling efficiency in complex inference tasks.

While the current model is trained on fixed-resolution simulations, recent advances in resolution-independent operator learning (Jiang et al., 2024) and discretization-agnostic Bayesian inference (Selig, 2014) also suggest promising directions for future work. Incorporating such methods could further improve model scalability and adaptability across different sensor designs and discretization levels.

Approximate Bayesian computation may also be an effective approach worth investigating to identify scalable alternatives, particularly in scenarios requiring faster inference or surrogate-based acceleration. Incorporating predictive uncertainty, for example, via Bayesian neural networks or ensemble-based approaches, would allow the likelihood function to account for input-dependent confidence levels, resulting in more reliable posterior estimates. This could also be an important direction to explore, which would likely improve the robustness of the inference framework, particularly in regions of the parameter space where the ML model exhibits lower confidence. These aspects, however, go well beyond the scope of this work and are left for future research.

CRedit authorship contribution statement

Ehsan Khodadadian: Visualization, Validation, Software, Investigation, Data curation. **Daniele Goldoni:** Writing – review & editing, Visualization, Validation, Methodology, Data curation, Software, Investigation, Conceptualization. **Jacopo Nicolini:** Writing – review & editing, Visualization, Data curation. **Amirreza Khodadadian:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Clemens Heitzinger:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Luca Selmi:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Frans Widdershoven (NXP Semiconductors) for providing the chips and for the fruitful discussion. D. Goldoni, J. Nicolini, and L. Selmi acknowledge partial support by the ESF REACT-EU National Operational Programme (PON) Research and Innovation 2014–2020, Austria, and by the European Union, NextGenerationEU initiative via Linea di Investimento 1.3 – Partenariati Estesi Area 6, PNRR HEAL ITALIA Project, PE 0000019. E. Khodadadian and A. Khodadadian acknowledge financial support from the Austrian Science Fund (FWF), Austria, Project no. P-36520, entitled *Using Single Atom Catalysts as Nanozymes in FET Sensors*.

Data availability

Data will be made available on request.

References

- Abbott, J., Mukherjee, A., Wu, W., Ye, T., Jung, H.S., Cheung, K.M., Gertner, R.S., Basan, M., Ham, D., Park, H., 2022. Multi-parametric functional imaging of cell cultures and tissues with a CMOS microelectrode array. *Lab Chip* 22, 1286–1296. <http://dx.doi.org/10.1039/D1LC00878A>.
- Antonelli, G., Filippi, J., D’Orazio, M., Curci, G., Casti, P., Mencattini, A., Martinelli, E., 2024. Integrating machine learning and biosensors in microfluidic devices: A review. *Biosens. Bioelectron.* 263, 116632. <http://dx.doi.org/10.1016/j.bios.2024.116632>, URL <https://www.sciencedirect.com/science/article/pii/S0956566324006389>.
- Asadian, E., Bahramian, F., Siavashy, S., Movahedi, S., Keçili, R., Hussain, C.M., Ghorbani-Bidkorpeh, F., 2024. A review on recent advances of AI-integrated microfluidics for analytical and bioanalytical applications. *TRAC Trends Anal. Chem.* 181, 118004. <http://dx.doi.org/10.1016/j.trac.2024.118004>, URL <https://www.sciencedirect.com/science/article/pii/S0165993624004874>.
- Atugoda, T., Piyumali, H., Wijesekara, H., Sonne, C., Lam, S.S., Mahatantila, K., Vithanage, M., 2023. Nanoplastic occurrence, transformation and toxicity: a review. *Environ. Chem. Lett.* <http://dx.doi.org/10.1007/s10311-022-01479-w>.
- Barsoukov, E., Macdonald, J.R., 2005. *Impedance Spectroscopy: Theory, Experiment, and Application*. Wiley-Interscience, <http://dx.doi.org/10.1002/0471716243>.
- Bhaiyya, M., Panigrahi, D., Rewatkar, P., Haick, H., 2024. Role of machine learning assisted biosensors in point-of-care-testing for clinical decisions. *ACS Sensors* 9 (9), 4495–4519. <http://dx.doi.org/10.1021/acssensors.4c01582>.
- Camerlingo, N., Siviero, I., Vettoretti, M., Sparacino, G., Del Favero, S., Facchinetti, A., 2023. Bayesian denoising algorithm dealing with colored, non-stationary noise in continuous glucose monitoring timeseries. *Front. Bioeng. Biotechnol.* 11, <http://dx.doi.org/10.3389/fbioe.2023.1280233>, URL <https://www.frontiersin.org/journals/bioengineering-and-biotechnology/articles/10.3389/fbioe.2023.1280233>.
- Carminati, M., Ferrari, G., Guagliardo, F., Sampietro, M., 2011. ZeptoFarad capacitance detection with a miniaturized CMOS current front-end for nanoscale sensors. *Sensors Actuators A: Phys.* 172 (1), 117–123. <http://dx.doi.org/10.1016/j.sna.2011.02.052>, URL <https://www.sciencedirect.com/science/article/pii/S0924424711001257>, EuroSensors XXIV, Linz, Austria, 5–8 September 2010.
- Cossetti, A., Brandalise, D., Palestri, P., Bertacchini, A., Ramponi, M., Widdershoven, F., Benini, L., Selmi, L., 2020. Ultra-high frequency (500 MHz) capacitance spectroscopy for nanobiosensing. In: 2020 IEEE SENSORS. pp. 1–4. <http://dx.doi.org/10.1109/SENSORS47125.2020.9278583>.
- Cossetti, A., Laborde, C., Brandalise, D., Widdershoven, F., Lemay, S.G., Selmi, L., 2021. Space and frequency dependence of nanocapacitor array sensors response to microparticles in electrolyte. *IEEE Sensors J.* 21 (4), <http://dx.doi.org/10.1109/JSEN.2020.3032712>.
- Cossetti, A., Selmi, L., 2018. On the response of nanoelectrode impedance spectroscopy measures to plant, animal, and human viruses. *IEEE Trans. Nanobioscience* 17 (2), <http://dx.doi.org/10.1109/TNB.2018.2826919>.
- Cossetti, A., Stadlbauer, B., Morales E., J.A., Taghizadeh, L., Selmi, L., Heitzinger, C., 2019. Determination of micro- and nano-particle properties by multi-frequency Bayesian methods and applications to nanoelectrode array sensors. In: 2019 IEEE SENSORS. pp. 1–4. <http://dx.doi.org/10.1109/SENSORS43011.2019.8956529>.
- Cui, F., Yue, Y., Zhang, Y., Zhang, Z., Zhou, H.S., 2020. Advancing biosensors with machine learning. *ACS Sensors* 5 (11), 3346–3364. <http://dx.doi.org/10.1021/acssensors.0c01424>.
- Eberhard, T., Casillas, G., Zarus, G.M., Barr, D.B., 2024. Systematic review of microplastics and nanoplastics in indoor and outdoor air: identifying a framework and data needs for quantifying human inhalation exposures. *J. Expo. Sci. Environ. Epidemiology* <http://dx.doi.org/10.1038/s41370-023-00634-x>.
- Faustmann, M., Melenk, J.M., Parvizi, M., 2022. *H*-matrix approximability of inverses of FEM matrices for the time-harmonic Maxwell equations. *Adv. Comput. Math.* 48 (5), 59.
- Feng, F., Kepler, T.B., 2015. Bayesian estimation of the active concentration and affinity constants using surface plasmon resonance technology. *PLoS One* 10 (6), 1–17. <http://dx.doi.org/10.1371/journal.pone.0130812>.
- Geremia, E., Muscarello Tomajoli, M.T., Murano, C., Petito, A., Fasciolo, G., 2023. The impact of micro- and nanoplastics on aquatic organisms: Mechanisms of oxidative stress and implications for human health—A review. *Environments* 10 (9), <http://dx.doi.org/10.3390/environments10090161>.
- Goldoni, D., Galstyan, V., Nappi, L., Rossella, F., D’Angelo, P., Rovati, L., Selmi, L., 2024a. Perspectives on multiparametric high-frequency impedance spectroscopy characterization of anatase TiO₂ nanotubes with nanoelectrode array sensors. *Meas.: Sensors* 101414. <http://dx.doi.org/10.1016/j.measen.2024.101414>.
- Goldoni, D., Ongaro, C., Orazi, L., Rovati, L., Selmi, L., 2023a. Estimation of analyte’s vertical positions above the surface of nanocapacitor array biosensors. In: 2023 IEEE SENSORS. IEEE, pp. 1–4. <http://dx.doi.org/10.1109/SENSORS56945.2023.10325169>.
- Goldoni, D., Rovati, L., Selmi, L., 2023b. Towards continuous nano-plastic monitoring in water by high frequency impedance measurement with nano-electrode arrays. *IEEE Sensors J.* <http://dx.doi.org/10.1109/JSEN.2023.3296158>, 1–1.
- Goldoni, D., Rovati, L., Selmi, L., 2024b. Size-resolved concentration estimation of nano- and micro-plastics for different water salinity with nanoelectrode array sensors. In: 2024 IEEE SENSORS. pp. 1–4. <http://dx.doi.org/10.1109/SENSORS60989.2024.10784490>.

- Green, P.J., Mira, A., 2001. Delayed rejection in reversible jump Metropolis–Hastings. *Biometrika* 88 (4), 1035–1053.
- Guillén, M.J., Nicolini, J., Goldoni, D., Madrid, R., Selmi, L., 2025. Electrical characterization of red blood cells with a nanoelectrode array sensor. In: *Proceedings of SIE 2024*. Springer Nature Switzerland, Cham, pp. 266–273. http://dx.doi.org/10.1007/978-3-031-71518-1_31.
- Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: efficient adaptive MCMC. *Stat. Comput.* 16, 339–354.
- Haario, H., Saksman, E., Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Statist.* 14, 375–395.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Hu, K., Arcadia, C.E., Rosenstein, J.K., 2021a. A large-scale multimodal CMOS biosensor array with 131,072 pixels and code-division multiplexed readout. *IEEE Solid-State Circuits Lett.* 4, 48–51. <http://dx.doi.org/10.1109/LSSC.2021.3056515>.
- Hu, K., Arcadia, C.E., Rosenstein, J.K., 2021b. Super-resolution electrochemical impedance imaging with a 100 × 100 CMOS sensor array. In: *2021 IEEE Biomedical Circuits and Systems Conference*. BioCAS, pp. 1–4. <http://dx.doi.org/10.1109/BioCAS49922.2021.9644965>.
- Huang, X., Wang, F., Zhang, B., Liu, H., 2024. Enriched physics-informed neural networks for dynamic Poisson–Nernst–Planck systems. *arXiv preprint arXiv:2402.01768*.
- Hunt-Smith, N.T., Melnitchouk, W., Ringer, F., Sato, N., Thomas, A.W., White, M.J., 2024. Accelerating Markov chain Monte Carlo sampling with diffusion models. *Comput. Phys. Comm.* 296, 109059.
- Jain, A., Chung, S., Spencer, E.A., Hall, D.A., 2024. An electrochemical CMOS biosensor array using phase-only modulation with 0.035% phase error and in-pixel averaging. *IEEE Trans. Biomed. Circuits Syst.* <http://dx.doi.org/10.1109/TBCAS.2024.3450843>.
- Jiang, X., Wang, X., Wen, Z., Wang, H., 2024. Resolution-independent generative models based on operator learning for physics-constrained Bayesian inverse problems. *Comput. Methods Appl. Mech. Engrg.* 420, 116690.
- Jung, D., Junek, G.V., Park, J.S., Kumashi, S.R., Wang, A., Li, S., Grijalva, S.I., Fernandez, N., Cho, H.C., Wang, H., 2021. A CMOS 21 952-pixel multi-modal cell-based biosensor with four-point impedance sensing for holistic cellular characterization. *IEEE J. Solid-State Circuits* 56 (8), 2438–2451. <http://dx.doi.org/10.1109/JSSC.2021.3085571>.
- Khodadadian, A., Parvizi, M., Heitzinger, C., 2020a. An adaptive multilevel Monte Carlo algorithm for the stochastic drift–diffusion–Poisson system. *Comput. Methods Appl. Mech. Engrg.* 368, 113163.
- Khodadadian, A., Parvizi, M., Teshnehlav, M., Heitzinger, C., 2022. Rational design of field-effect sensors using partial differential equations, Bayesian inversion, and artificial neural networks. *Sensors* 22 (13), 4785.
- Khodadadian, A., Stadlbauer, B., Heitzinger, C., 2020b. Bayesian inversion for nanowire field-effect sensors. *J. Comput. Electron.* 19, 147–159.
- Kumashi, S.R., Jung, D., Park, J., Sanz, S.T., Grijalva, S., Wang, A., Li, S., Cho, H.C., Ajo-Franklin, C.M., Wang, H., 2021. A CMOS multi-modal electrochemical and impedance cellular sensing array for massively paralleled exoelectrogen screening. *IEEE Trans. Biomed. Circuits Syst.* 15 (2), 221–234. <http://dx.doi.org/10.1109/TBCAS.2021.3068710>.
- Laborde, C., Pittino, F., Verhoeven, H.A., Lemay, S.G., Selmi, L., Jongmsa, M.A., Widdershoven, F.P., 2015. Real-time imaging of microparticles and living cells with CMOS nanocapacitor arrays. *Nature Nanotechnology* 10 (9), <http://dx.doi.org/10.1038/nnano.2015.163>.
- Lai, L.-H., Lin, W.-Y., Lu, Y.-W., Lui, H.-Y., Yoshida, S., Chiou, S.-H., Lee, C.-Y., 2023a. A 460 800 pixels CMOS capacitive sensor array with programmable fusion pixels and noise canceling for life science applications. *IEEE Trans. Circuits Syst. II: Express Briefs* 70 (5), 1734–1738. <http://dx.doi.org/10.1109/TCSII.2023.3261445>.
- Lai, P.-H., Tseng, L.-S., Yang, C.-M., Lu, M.S.C., 2023b. Design and characterization of a 16 × 16 CMOS capacitive DNA sensor array. *IEEE Sensors J.* 23 (8), 8120–8127. <http://dx.doi.org/10.1109/JSEN.2023.3253123>.
- Lee, J., Chae, K.J., 2021. A systematic protocol of microplastics analysis from their identification to quantification in water environment: A comprehensive review. <http://dx.doi.org/10.1016/j.jhazmat.2020.124049>.
- Lee, D., Jung, D., Jiang, F., Junek, G.V., Park, J., Liu, H., Kong, Y., Wang, A., Kim, Y., Choi, K.-S., Wang, J., Wang, H., 2023. A multi-functional CMOS biosensor array with on-chip DEP-assisted sensing for rapid low-concentration analyte detection and close-loop particle manipulation with no external electrodes. *IEEE Trans. Biomed. Circuits Syst.* 17 (6), 1214–1226. <http://dx.doi.org/10.1109/TBCAS.2023.3343068>.
- Lemay, S.G., Laborde, C., Renault, C., Cossetini, A., Selmi, L., Widdershoven, F.P., 2016. High-frequency nanocapacitor arrays: Concept, recent developments, and outlook. *Acc. Chem. Res.* 49 (10), 2355–2362. <http://dx.doi.org/10.1021/acs.accounts.6b00349>.
- Lombardo, F., Pittino, F., Goldoni, D., Selmi, L., 2024. Machine learning and data augmentation methods for multispectral capacitance images of nanoparticles with nanoelectrodes array biosensors. *Eng. Appl. Artif. Intell.* 127, 107246. <http://dx.doi.org/10.1016/j.engappai.2023.107246>.
- Lopez, C.M., Chun, H.S., Berti, L., Wang, S., Putzeys, J., Van Den Bulcke, C., Weijers, J.-W., Firrincieli, A., Reumers, V., Braeken, D., Van Helleputte, N., 2018. A 16384-electrode 1024-channel multimodal CMOS MEA for high-throughput intracellular action potential measurements and impedance spectroscopy in drug-screening applications. In: *2018 IEEE International Solid-State Circuits Conference - ISSCC*, pp. 464–466. <http://dx.doi.org/10.1109/ISSCC.2018.8310385>.
- Marin, J.-M., Pudlo, P., Robert, C.P., Ryder, R.J., 2012. Approximate Bayesian computational methods. *Stat. Comput.* 22 (6), 1167–1180.
- Miccoli, B., Lopez, C.M., Goikoetxea, E., Putzeys, J., Sekeri, M., Krylychkina, O., Chang, S.-W., Firrincieli, A., Andrei, A., Reumers, V., Braeken, D., 2019. High-density electrical recording and impedance imaging with a multi-modal CMOS multi-electrode array chip. *Front. Neurosci.* 13, <http://dx.doi.org/10.3389/fnins.2019.00641>.
- Mirsian, S., Hilber, W., Khodadadian, E., Parvizi, M., Khodadadian, A., Khoshfetrat, S.M., Heitzinger, C., Jakoby, B., 2025. Graphene-based FETs for advanced biocatalytic profiling: investigating heme peroxidase activity with machine learning insights. *Microchim. Acta* 192 (3), 199.
- Noii, N., Khodadadian, A., Ulloa, J., Aldakheel, F., Wick, T., Francois, S., Wriggers, P., 2022. Bayesian inversion with open-source codes for various one-dimensional model problems in computational mechanics. *Arch. Comput. Methods Eng.* 29 (6), 4285–4318.
- Nøvik, S., Drageseth, M.F., Grøndalen, M.B., Nilsen, O., Krauss, S.J.K., Martinsen, O.G., Hafliger, P.D., 2022. A CMOS multi-electrode array for four-electrode bioimpedance measurements. *IEEE Trans. Biomed. Circuits Syst.* 16 (6), 1276–1286. <http://dx.doi.org/10.1109/TBCAS.2022.3214243>.
- Osman, A.I., Hosny, M., Eltaewil, A.S., Omar, S., Elgarahy, A.M., Farghali, M., Yap, P.-S., Wu, Y.-S., Nagandran, S., Batumalaie, K., Gopinath, S.C.B., John, O.D., Sekar, M., Saikia, T., Karunanithi, P., Hatta, M.H.M., Akinyede, K.A., 2023. Microplastic sources, formation, toxicity and remediation: a review. *Environ. Chem. Lett.* <http://dx.doi.org/10.1007/s10311-023-01593-3>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pittino, F., 2015. Prospects of Nanoelectronic Biosensing with High-Frequency Impedance Spectroscopy (Ph.D. thesis).
- Pittino, F., Scarbolo, P., Widdershoven, F., Selmi, L., 2015. Derivation and Numerical Verification of a Compact Analytical Model for the AC Admittance Response of Nanoelectrodes, Suitable for the Analysis and Optimization of Impedance Biosensors. *IEEE Trans. Nanotechnol.* 14 (4), <http://dx.doi.org/10.1109/TNANO.2015.2434106>.
- Pittino, F., Selmi, L., 2014. Use and comparative assessment of the CVFEM method for Poisson–Boltzmann and Poisson–Nernst–Planck three dimensional simulations of impedimetric nano-biosensors operated in the DC and AC small signal regimes. *Comput. Methods Appl. Mech. Engrg.* 278, <http://dx.doi.org/10.1016/j.cma.2014.06.006>.
- Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707.
- Sajjad, M., Huang, Q., Khan, S., Khan, M.A., Liu, Y., Wang, J., Lian, F., Wang, Q., Guo, G., 2022. Microplastics in the soil environment: A critical review. *Environ. Technol. Innov.* 27, 102408. <http://dx.doi.org/10.1016/j.eti.2022.102408>.
- Sangkham, S., Faikhaw, O., Munkong, N., Sakunkoo, P., Arunlertaree, C., Chavali, M., Mousazadeh, M., Tiwari, A., 2022. A review on microplastics and nanoplastics in the environment: Their occurrence, exposure routes, toxic studies, and potential effects on human health. <http://dx.doi.org/10.1016/j.marpolbul.2022.113832>.
- Schoeberl, J., 1997. NETGEN An advancing front 2D/3D-mesh generator based on abstract rules. *Comput. Vis. Sci.* 1, 41–52. <http://dx.doi.org/10.1007/s007910050004>.
- Selig, M., 2014. The NIFTY way of Bayesian signal inference. In: *AIP Conference Proceedings*. Vol. 1636, American Institute of Physics, pp. 68–73.
- Smith, R., 2013. Uncertainty Quantification: Theory, Implementation, and Applications. Vol. 12, SIAM.
- Stadlbauer, B., Cossetini, A., Morales E., J.A., Pasterk, D., Scarbolo, P., Taghizadeh, L., Heitzinger, C., Selmi, L., 2019. Bayesian estimation of physical and geometrical parameters for nanocapacitor array biosensors. *J. Comput. Phys.* 397, <http://dx.doi.org/10.1016/j.jcp.2019.108874>.
- Sunnåker, M., Busetto, A.G., Numminen, E., Corander, J., Foll, M., Dessimoz, C., 2013. Approximate Bayesian computation. *PLoS Comput. Biol.* 9 (1), e1002803.
- Tabrizi, H.O., Forouhi, S., Ghafar-Zadeh, E., 2022. A high dynamic range dual 8 × 16 capacitive sensor array for life science applications. *IEEE Trans. Biomed. Circuits Syst.* 16 (6), 1191–1203. <http://dx.doi.org/10.1109/TBCAS.2022.3230668>.
- Taghizadeh, L., Khodadadian, A., Heitzinger, C., 2017. The optimal multilevel Monte-Carlo approximation of the stochastic drift–diffusion–Poisson system. *Comput. Methods Appl. Mech. Engrg.* 318, 739–761.

- Tran, T.V., Jilil, A., Nguyen, T.M., Nguyen, T.T.T., Nabgan, W., Nguyen, D.T.C., 2023. A review on the occurrence, analytical methods, and impact of microplastics in the environment. *Environ Toxicol Pharmacol* 102, 104248. <http://dx.doi.org/10.1016/j.etap.2023.104248>.
- Wang, S., Perdikaris, P., 2021. Learning the solution operator of parametric partial differential equations with physics-informed DeepONets. *Sci. Adv.* 7 (40), eabi8605.
- Widdershoven, F., 2025. Pixelated capacitive sensors (PCS) for embedded multi-sensing. In: Makinwa, K.A.A., Baschiroto, A., Nauta, B. (Eds.), *Imaging Sensors, Power Management, PLLs and Frequency Synthesizers: Advances in Analog Circuit Design 2023*. Springer Nature Switzerland, Cham, pp. 23–35. http://dx.doi.org/10.1007/978-3-031-71559-4_2.
- Widdershoven, F., Cossettini, A., Laborde, C., Bandiziol, A., Van Swinderen, P.P., Lemay, S.G., Selmi, L., 2018. A CMOS pixelated nanocapacitor biosensor platform for high-frequency impedance spectroscopy and imaging. *IEEE Trans. Biomed. Circuits Syst.* 12 (6), <http://dx.doi.org/10.1109/TBCAS.2018.2861558>.
- Widdershoven, F., Van Steenwinkel, D., Überfeld, J., Merelle, T., Suy, H., Jedema, F., Hoofman, R., Tak, C., Sedzin, A., Cobelens, B., Sterckx, E., Van Der Werf, R., Verheyden, K., Kengen, M., Swartjes, F., Frederix, F., 2010. CMOS biosensor platform. In: *Technical Digest - International Electron Devices Meeting, IEDM*. <http://dx.doi.org/10.1109/IEDM.2010.5703484>.
- Wiese, J.G., Wimmer, L., Papamarkou, T., Bischl, B., Günemann, S., Rügamer, D., 2023. Towards efficient MCMC sampling in Bayesian neural networks by exploiting symmetry. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 459–474.
- Yusuf, A., Sodiq, A., Giwa, A., Eke, J., Pikuda, O., Eniola, J.O., Ajiwokewu, B., Sambudi, N.S., Bilad, M.R., 2022. Updated review on microplastics in water, their occurrence, detection, measurement, environmental pollution, and the need for regulatory standards. <http://dx.doi.org/10.1016/j.envpol.2021.118421>.
- Zhou, Z., Tian, D., Yang, Y., Cui, H., Li, Y., Ren, S., Han, T., Gao, Z., 2024. Machine learning assisted biosensing technology: An emerging powerful tool for improving the intelligence of food safety detection. *Curr. Res. Food Sci.* 8, 100679. <http://dx.doi.org/10.1016/j.crfs.2024.100679>, URL <https://www.sciencedirect.com/science/article/pii/S2665927124000054>.