MIcroelectornics RELiability driven by Artificial Intelligence
Project no. 101072491

# Deliverable 6.1

# Implementation strategy for weakest link physics of degradation

WP 6 – AI-based reliability

Version 01

| | |
|---|---|
| **Authors** | René Poelma (Nexperia), Willem van Driel (TU Delft), Stefano Mariani (POLIMI), Shehryar Imtiaz (DC12, POLIMI), Zihan Zhang (DC5, TU Delft) |
| **Lead participant** | Nexperia |
| **Delivery date** | 31/05/2024 |
| **Dissemination level** | Public |
| **Type** | Report |

Funded by the European Union

Engineering and Physical Sciences Research Council

Funded by the European Union (Grant Agreement No. 101072491). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.
Partners from the UK are supported by the UK Engineering and Physical Sciences Research Council.

# Revision history

| Author(s) | Description | Date |
|---|---|---|
| Shehryar Imtiaz (DC12), Zihan Zhang (DC5) | Draft deliverable | 11.05.2024 |
| René Poelma, Willem van Driel, Stefano Mariani, Shehryar Imtiaz, Zihan Zhang | First review | 18.05.2024 |
| Shehryar Imtiaz, Zihan Zhang | Final version | 31.05.2024 |

# Executive summary

### Background

Deliverable D6.1 is part of the Work Package (WP) 6 - AI-based reliability. In this WP, we will introduce physics of degradation-informed machine learning, making use of the results of WP4 and WP5 to predict and improve the electronic system product reliability.

### Objectives

Deep learning-based interpretation of measurement data and derivation of yet not understood interrelations. Integration and implementation of (noisy) data and mathematical models through neural networks or other kernel-based regression networks. Design of specialized network architectures that automatically satisfy some of the physical invariants for better accuracy, faster training and improved generalization for electronic systems.

### Methodology and implementation

The main method will be based on machine learning technique and physics implementation machine learning, the specific application will rely on each DC's work.

### Outcomes

Recent advancements in AI processors have significantly enhanced computational efficiency, enabling the development of sophisticated neural networks. This research introduces a novel AI-based structural analysis method using Physics-Informed Neural Networks (PINNs) to monitor weak link physics in microelectronics, incorporating governing equations and boundary conditions into the training process. Validated against analytical solutions and finite element methods (FEM), these innovative techniques hold promise for revolutionising weak link physics analysis and potentially other engineering applications.

### Impact

The objectives for WP6 are:

- Physics-informed machine learning modelling strategy for electric components
- Global criticality assessment tool for defined features in a PCBA
- Prediction of the remaining useful lifetime of electronic systems by the use of machine learning.

By completing D6.1, the third objective is satisfied.

### Next steps

In the next steps, the methods will be employed into each DC's own work in the propose to complete WP6.

# Contents

# Partner short names

| Short name | Partner name |
|---|---|
| PCCL | Polymer Competence Center Leoben |
| POLIMI | Politecnico di Milano |
| TU Chemnitz | Technische Universität Chemnitz |
| TU Delft | Technische Universiteit Delft |
| IMEC | Interuniversitair Micro-Electronica Centrum |
| ams OSRAM | Ams-Osram AG |
| AT&S | AT & S Austria Technologie & Systemtechnik Aktiengesellschaft |
| Bosch | Robert Bosch GmbH |
| Nanotest | Berliner Nanotest und Design GmbH |
| Nexperia | Nexperia BV |
| NXP | NXP Semiconductors Netherlands BV |
| SISW | Siemens Industry Software NV |
| Technoprobe | Technoprobe SPA |
| UOG | University of Greenwich |
| MCS | Materials Consultancy Services Limited |
| accelCH | accelopment Schweiz AG |
| KU Leuven | Katholieke Universiteit Leuven |
| MUL | Montanuniversität Leoben |
| MCL | Materials Center Leoben Forschungs GmbH |
| signify | Signify Netherlands BV |

# Abbreviations

| Acronym | Full name |
|---|---|
| **AI** | Artificial intelligence |
| **ANN** | Artificial Neural Network |
| **BPNN** | Backpropagation Neural Network |
| **D** | Deliverable |
| **DBCCB** | Direct Bond Copper |
| **DNN** | Deep Neural Networks |
| **DoS** | Design-on-Simulation |
| **EC** | European Commission |
| **EU** | European Union |
| **FEM** | Finite Element Methods |
| **GaN** | Gallium Nitride |
| **HEU** | Horizon Europe |
| **IGBT** | Insulated-Gate Bipolar Transistor |
| **M** | Month |
| **MS** | Milestone |
| **PDE** | Partial Differential Equation |
| **PINN** | Physics-Informed Neural Network |
| **RF** | Random Forests |
| **RNN** | Recurrent Neural Network |
| **RUL** | Remaining Useful Lifetime |
| **SVR** | Support Vector Regression |
| **Tj** | Junction Temperature |
| **WP** | Work Package |
| **3D** | Three-dimensional |

# 1 Introduction

**Note: This document will be reviewed every year, for new concepts, results and methodologies!**

The weak link physics represents the part of a device that is the first to cause system degradation or even failure. Devices in different operating conditions may have different weak link physics and correspond to different failure modes. For example, devices in ships tend to exhibit failure modes caused by humid environments, devices in aircraft exhibit shock-induced failures, and so on. Table 1 shows some of the failure modes of insulated-gate bipolar transistor (IGBT) modules and the corresponding causes. Note that even though the focus of research is now shifting to gallium nitride and silicon carbide, the packaging of these new devices is still not significantly different, so many of the major failure modes are similar to those of IGBTs.

*Table 1: Failure modes and causes for IGBT module [1]*

| Failure Type | Cause |
|---|---|
| Thermal-mechanical Failure | Bonding wire, die-attach, delamination, passivation cracks, etc. |
| Thermal Failure | Package degradation; Cooling system Faults |
| Corrosion | Chemical reactions due to humidity; Flashover due to formation of conductive parts |
| Overcurrent | Unexpected electric load; Improper control action |
| Gate Overvoltage | Gate driver anomalies; External surges; Gate driver anomalies |
| Collector emitter Overvoltage | Unexpected electric load; Improper control signal |
| Manufacturing Defects | Bond wire defects; Solder defects; Metallization defects; DBC Insulation defects |

The first step in the implementation is to find the weak physical link that causes a specific failure mode, either through accelerated experimentation or simulation. Usually, thermal cycling or power cycling experiments are used, and the simulation approach is similar. It is worth mentioning that the failure modes in the laboratory are often different from those in real applications due to unknown operating conditions.

The second step is to characterise the lifetime model for a specific failure mode, and a summary of some of the approaches is given in Table 2. Some short descriptions for the lifetime model are shown in Table 3.

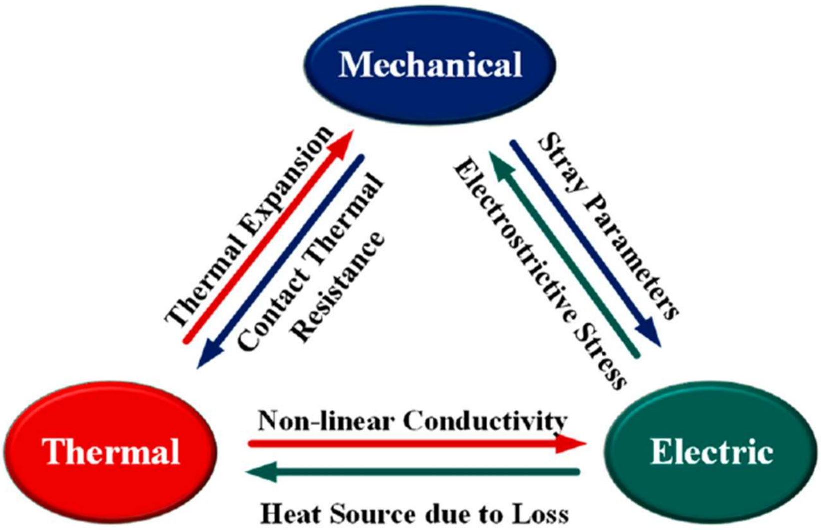
*Table 2: Failure modes and corresponding lifetime models [1]*

| Failure mechanisms | Failure modes | Lifetime Model |
|---|---|---|
| Fatigue | Bonding wire cracks, lift-off | Empirical model[2] |
| Surface Fatigue | Metallization reconstruction | Empirical model[2] |
| Solder Joint Fatigue | Solder layer cracks, delamination | Models based on Coffin-Manson[2]; Empirical as Norris-Landzberg[3] |
| Ceramic Fatigue | Direct Bond Copper substrate cracks | Empirical[2] |
| Electro-Migration | Bonding wire metallization, void formation | Empirical as Black's law[2] |

Table3: Lifetime Model description.

| Model | Description |
|---|---|
| Coffin-Manson model | Predicts fatigue life based on plastic strain amplitudes and cycles to failure. |
| Norris-Landzberg model | Extends Coffin-Manson by including temperature and frequency effects for solder joints. |
| Empirical model for Bonding wire cracks and lift-off | Predicts bonding wire failure due to thermal cycling, mechanical stress, and material properties. |
| Empirical model for Metallization reconstruction | Estimates metallization layer lifespan considering current density and temperature. |
| Empirical model for DCB substrate cracks | Predicts DCB substrate cracks due to thermal cycling and mechanical stress. |
| Empirical model based on Black's law for Electro-Migration | Uses Black's law to predict electro-migration failure based on current density and temperature. |

The third step is the practical application of the life models. The life model is essentially a fitted function, where the input is the measured data from the sensor and the output is the current remaining life. Typical input data include junction temperature (Tj), Vce in IGBT module, etc. Thermal, electrical, and mechanical parameters all change and couple with each other as the degradation process occurs (see Fig. 1), and the researcher needs to select the most suitable parameter according to the specific failure modes and the available sensors. As mentioned before, due to the different actual working conditions, the life models in applications actually need to be updated in real-time, which can be achieved by digital twin technology or machine learning based methods. In simple terms, the digital twin model is a real-time updated data model that contains not only the real-time lifetime model but also various current state parameters of the device, a review of which can be found in the previous deliverables. In this deliverable, we will discuss how different strategies can be used to extract and monitor weak link physics in a device and perform health diagnostics using local data measured by sensors. Also, the definition of different level of digital twin is a digital twin is a comprehensive simulation that integrates multiphysics, multiscale, and probabilistic models to replicate the lifecycle of a physical system. The definition and utility of a digital twin depend on the level of detail and accuracy of its simulation. At lower accuracy levels, it is characterised as a three-dimensional model consisting of physical elements, simulations or virtual twins, and their interconnections. In contrast, a five-dimensional digital twin, with higher accuracy, enhances the precision of the simulation.




*Direct Bond Copper*

# 2 Machine learning applications on monitoring weakest link physics

## 2.1 Introduction

The weakest link physics focuses on identifying and understanding the most vulnerable parts of a device that are likely to fail first under operational stresses (physical stress, electrical stress or thermal stress). This is crucial for monitoring the reliability of the devices, where a lot of data will be generated during the operation from various conditions and devices which is desirable for a machine learning algorithm.

Machine learning has shown promise in areas such as image recognition and parameter fitting due to its superior nonlinear fitting capabilities. By applying these techniques to monitor the weakest link physics, researchers can predict the location and number of defects, optimise sensor placements, and improve the accuracy of lifetime predictions with more input variables. This involves using various machine learning models, including classical neural networks, Support Vector Regression (SVR), Random Forests (RF), and Physics-Informed Neural Networks (PINNs), to analyse real-time data and update models in real-time.

## 2.2 Data-driven machine learning

Traditional data-driven machine learning methods can be well applied in the fields of image recognition and parameter fitting by virtue of their superior nonlinear fitting performance. In this chapter, the possibilities of machine learning for weak-link physics in monitoring devices are discussed.

In [5], a machine learning model is used to predict the distribution of defects in the solder layer in a single-side cooled module, and therefore, the weak-link physics in this case is defined as the solder layer. The researcher divided the solder layer into a finite number of units and predicted the location and number of defects occurring in the solder had by a classical neural network with two hidden layers, which was obtained from the results of thermal simulation. The researchers found that the use of stress parameters to characterise the defects gives better predictions, while the number and location of sensors are also critical to the predicted results. The model enables prediction of failure in weak-link physics (solder layers) in the device, which makes prediction of remaining life and pre-diagnosis possible as well.

The literature indicates that the use of stress as an indicator of weak-link physics is better, collectively, most studies tend to favour the use of thermal indicators to characterise the degradation process. In [6], an on-die machine learning model is used to monitor a GaN device in real time. The model is able to update the thermal model of the device based on real-time data without interrupting the operation of the device, thus monitoring the junction temperature in real time. If the junction temperature is known, then it can be used to predict the remaining lifetime of the device, and this machine learning approach greatly increases the accuracy of the model over time. Similarly, some researchers have combined traditional thermal network models with machine learning methods to build self-tuning thermal models [7]. The model adjusts the parameters in the thermal network based on real-time temperature inputs, thus achieving the requirement of updating the model in real time and accurately predicting the junction temperature.

Another implementation strategy is the pre-modelling of weak link physics. For example, in [8], the junction temperature data of IGBTs at different peak currents, switching frequencies, and ambient temperatures were fitted using three machine learning methods using simulation data, including SVR, RF, and BPNN. The results show that both SVR and BPNN models are able to predict the junction temperatures of the IGBT module within acceptable errors, and therefore can be used as a weak link physics model for the module in real-world conditions. module's weak link physics monitoring methodology can be applied in real-world operating conditions.

# 3 Use of Artificial Intelligence models to oversee and mitigate the weak link physics.

## 3.1 Introduction

Probe cards are the building blocks of any chip that is used in the manufacturing of every electronic device present in this day and age. Several performance indicators affect the reliability of these probe cards. These indicators are upper and lower pad sizes, solder volume, buffer layer thickness, and chip thickness, etc. Normally, the accelerated thermal cycling test is used to evaluate the reliability life of a probe card; however, optimising the design parameters through this technique is time-consuming, expensive, and most importantly, reducing the number of experiments becomes a critical point. In recent years, many studies have adapted the finite-element-based Design-on-Simulation (DoS) technology for the reliability assessment of these probe cards. This technology can effectively reduce the design cycle, reduce costs, and effectively optimise the reliability of a probe card. However, the simulation analysis results are highly dependent on the individual researcher and are usually inconsistent between them. Artificial intelligence (AI) can help researchers avoid the shortcomings of the human factor. This research work intends to develop an AI-assisted model by combining artificial intelligence and simulation technologies to predict probe card key performance indicators devising a strategy for weak link physics of degradation. In order to ensure reliability prediction accuracy, the simulation procedure will be validated by several experiments prior to creating a large AI training database. This research will explore and implement several machine learning models, including Artificial Neural Network (ANN), Recurrent Neural Network (RNN) and Deep Neural Networks (DNN). An important fact is that in the area of artificial intelligence and machine learning, the implementation or building of an artificial intelligence is a less tedious task than the formulation of a dataset. A dataset can be of different modes, types, extensions, sizes and most importantly diversity. So, this is one of the areas that can be really exciting in this research topic. Intention is to develop an artificial intelligence model that can take into account not only the accuracy of our predictions but at the same time, it can also handle all the above-mentioned traits of any dataset.

## 3.2 Physics Informed Neural Network (PINN)

The first step in this endeavour is to use Physics Informed Neural Network framework to analyse the nonlinear buckling behaviour of a three-dimensional (3D) FG porous, slender beam resting on a Winkler-Pasternak foundation. The structure of the Physics Informed Neural Network can be seen in Fig.2. The alloys with which these beams are made of are devoid of any closed form solutions because of having different orders of Partial Differential Equations (PDEs) considering they are a mixture of various metals. Furthermore, each metal of the alloy has a non-linearity characteristic attached to it which makes the whole process extremely tedious, even if the PDEs does exist for these metals. These alloys/heterogenous materials are also called FG materials or functionally graded materials.

In order to avoid the complex process of Finite Element Modelling of PDEs, Physics Informed Neural Network is used here which insert the PDE and the boundary conditions into the loss function which is than trained by a gradient based optimiser. Two types of losses are fundamentally incorporated in PINNs on which the network is trained. One is the conventional loss which is the sum of three subsequent losses like loss of the governing PDE, loss for the boundary condition and lastly loss for the initial condition, while the other loss is the energy-based loss which is calculated by the partial derivatives of the outputs with respect to the inputs. This is achieved by the libraries of TensorFlow and Pytorch.

Couple of features that makes PINNs so efficient is the use of an Automatic Differentiator which basically simplifies the process of deep learning coding and the use of DeepXDE library. What happens is that the buckling of the beam is a 4th order PDE, and this library has dedicated modules that lets you define computational domain, PDEs, boundary and initial conditions, architecture of your neural network, constraints, training data and training hyperparameters. The beam theory is than incorporated in the architecture of the Physics Informed Neural Network framework. The total potential energy is considered as a loss and then minimised following the potential energy principle. For the purpose of determining the nonlinear bucking load, PINNs takes five inputs and outputs the critical buckling load. These inputs are length of the beam, width of the beam, thickness of the beam, beam's young's modulus and finally the porosity factor of the beam. If any more input parameters are taken as inputs than this will only make the prediction computationally intensive and expensive as well, which we are trying to avoid in the first place. This network is fundamentally made up of six hidden layers with 150 training points and a specific of number of iterations. Adam Optimiser is used here to conduct the training procedure to make the process more robust.

## 3.3 Implementation Strategy

Based on the beam theory, the first buckling load is usually determined by finding the lowest total potential energy. Therefore, we will use the energy-based model to compute it. Additionally, the PINN algorithm will take five input parameters as shown in figure 2, including l, b, h, E, and outputs the critical buckling load. The increase in input parameters might result in more computationally expensive predictions and harder to converge. Other parameters will be neglected as they are constant in the dataset. We will choose all the hyperparameters by trial and error in light of the accuracy and convergence of the solutions and computational efficiency.

It is observed in light of different implementations that a larger learning rate results in inaccurate solutions; conversely, a higher number of hidden layers results in the accuracy of the network's predictions at the cost of computational time. Greater than six hidden layers generates identical predictions, whereas one gives an inaccurate solution for a smaller number of hidden layers. Consequently, we intend to design a six-hidden layer network with the energy-based loss function. The PINN algorithm is constructed to find the global minimum that corresponds to the buckling response. Therefore, we can make sure that the network converges after training.

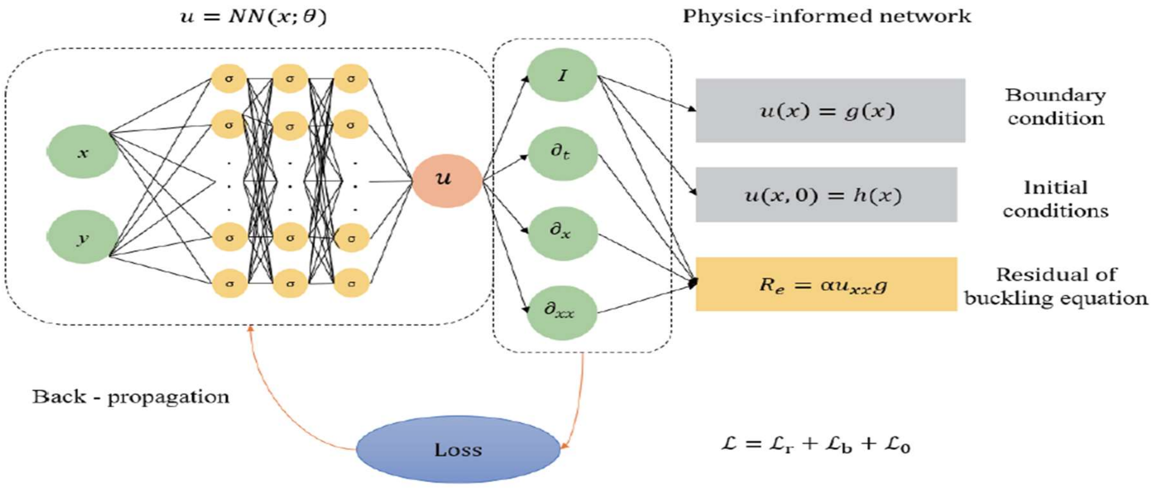



*Figure 2: Classic structure of PINN.*

## 3.4 Results

PINNs show extremely encouraging results predicting a near perfect nonlinear buckling load. This Nonlinear buckling load would be critical in the estimation of the physics of degradation of a probe card pin. This can further give us a clear pathway ahead to change the strategy once the number of geometric parameters to be considered also increase. The results can be summarized as follows (Fig.3):

### 3.4.1 Interpretation Of Results

- From Fig.3a and Fig.3b we can deduce that if we increase the Material Constant, Axial Functionally Graded Index and Thickness FG index, then the nonlinear buckling load increases.
- Fig.3c tells us that if we increase the Width FG Index and the Length Thickness, then the nonlinear buckling load decreases.
- Fig.3d shows a steady decrease of loss with every epoch which shows the effectiveness of the loss function and at a later stage, the activation function as well.

Due to the factual approximation accuracy, there is huge scope for this network to be used extensively. It can be an efficient tool for the approximation of various mechanical properties of different material such as Gold, Silver etc. but in my opinion this network can be extremely fruitful if used to gauge and predict the deformation of different materials. This is the reason that that this network can be incorporated to see the buckling effect of probe pins on a probe card as the common denominator between these two problems can be the nonlinear behaviour of both the FG materials and the probe pins of a probe card. In both cases the geometric nonlinearity of the beam and the Probe Pin is taken into account in a neural network and Artificial Intelligence based algorithm is used to predict performance parameters.

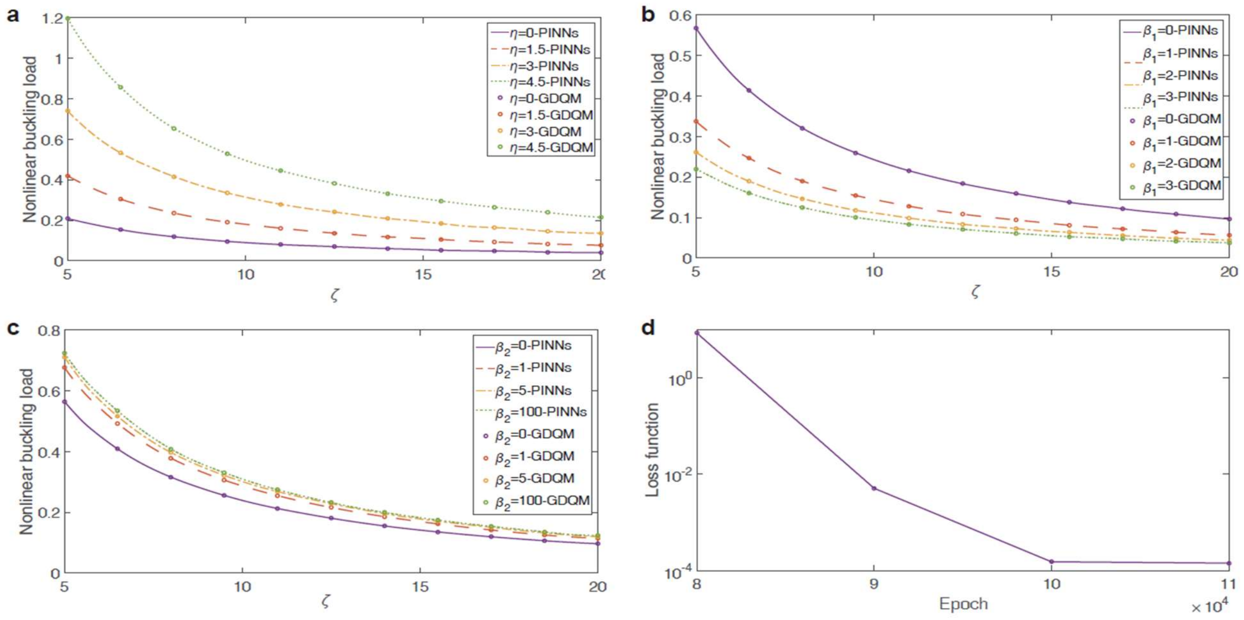



*Figure 3: a, b, c). Nonlinear buckling load prediction by PINN. d). Loss after each epoch.*

# 4 Conclusion and Future work

The computational efficiency of artificial intelligence has been exceptionally improved in recent years, with the availability of fast AI processors tail-made for massive neural networks. To facilitate the adoption of new AI technology in microelectronics, this research proposes an innovative artificial intelligence-based structural analysis method to oversee the weak link physics through Physics Informed Neural Network. The governing equations for the problems are discussed and incorporated into the physics-informed neural networks. The physical information provided by the governing equations and boundary conditions is used to orientate the training process creating an unsupervised learning procedure. The PINN framework and the training procedure are provided, where an adaptive loss weight control algorithm and the transfer learning technique are proposed to improve numerical efficiency.

These new methods have been thoughtfully validated by well-established methods in terms of the analytical solutions and the FEM. As the potential to be the next-generation weak link physics analysis method, this research innovatively adopts the emerging computational technique based on machine learning and artificial intelligence to solve second-order analysis problems for single members that could be motivated for other engineering problems.

# 5 References

[1] A. Abuelnaga, M. Narimani and A. S. Bahman, “A Review on IGBT Module Failure Modes and Lifetime Testing,” in IEEE Access, vol. 9, pp. 9643-9663, 2021, doi: 10.1109/ACCESS.2021.3049738.

[2] L. A. Escobar and W. Q. Meeker, “A review of accelerated test models,” Stat. Sci., vol. 21, no. 4, pp. 552–577, Nov. 2006.

[3] A. Volke and M. Hornkamp, IGBT Modules: Technologies, Driver and Application. Neubiberg, Germany: Infineon Technologies AG, 2017.

[4] Y. Wang, Y. Ding, Y. Yin, “Reliability of Wide Band Gap Power Electronic Semiconductor and Packaging: A Review,” Energies. 2022; 15(18):6670. https://doi.org/10.3390/en15186670

[5] M. Usui, M. Kuwahara, R. Okachi and J. Muramatsu, "Prediction of Defect Distribution in Solder Joint Using Device Stress Toward Prognostic Health Management of Power Modules," in IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 13, no. 9, pp. 1344-1352, Sept. 2023, doi: 10.1109/TCPMT.2023.3311321.

[6] Y. Huang, Y. Chen and D. B. Ma, "Enabling Online GaN Power Device Self-Health Monitoring With Analog SGD Supervised Learning and Tj-Independent Precursor Measurement," in IEEE Journal of Solid-State Circuits, doi: 10.1109/JSSC.2023.3330835.

[7] K. Rönnberg, P. Kakosimos, Z. Kolondjovski and E. Nordlund, "Machine learning-based adjustments of thermal networks," 11th International Conference on Power Electronics, Machines and Drives (PEMD 2022), Hybrid Conference, Newcastle, UK, 2022, pp. 424-428, doi: 10.1049/icp.2022.1087.

[8] J. Miao, Q. Yin, H. Wang, Y. Liu, H. Li and S. Duan, "IGBT Junction Temperature Estimation Based on Machine Learning Method," 2020 IEEE 9th International Power Electronics and Motion Control Conference (IPEMC2020-ECCE Asia), Nanjing, China, 2020, pp. 1-5, doi: 10.1109/IPEMC-ECCEAsia48364.2020.9395640.