# Microelectronics Failure Prediction and Localization from Optical and Thermal Imagery

Md Alimoor Reza*, Manjulata Garimella†, Sami Alajouni‡, Kerry Maize‡,
Peter Bermel‡, David Crandall† and Ali Shakouri‡
* Dept. of Mathematics and Computer Science, Drake University, Des Moines, Iowa md.reza@drake.edu
† Dept. of Computer Science, Indiana University, Bloomington, Indiana {mchivuku, djcran}@iu.edu
‡ Dept. of Electrical Engineering, Purdue University, West Lafayette, Indiana{salajlou, kmaize, pbermel, shakouri}@purdue.edu

***Abstract*—Understanding and predicting microelectronic failures is important for ensuring the reliability of modern electronic devices. In this paper, we develop a set of computer vision algorithms for modeling device failure due to electromigration, in which metal atoms are displaced due to current flow. The experimental setup contains a series of optical and thermal images of aluminum interconnects. We propose deep neural networks for two specific problems: predicting where in the device a failure will occur, and estimating the remaining lifespan before failure. We pose the former as a segmentation problem and solve it with a convolutional neural network (CNN) trained on multi-scale optical images. We pose the latter as a regression problem for which we designed a convolutional neural network augmented with a recurrent module (RNN) to model the temporal dimension. We compare against two baseline networks, finding that our model can predict the age of the aluminum more accurately from the optical images instead of the thermal ones. This work is the first that pursues a deep neural network-based modeling to predict electromigration failures from experimental optical and thermal images.**

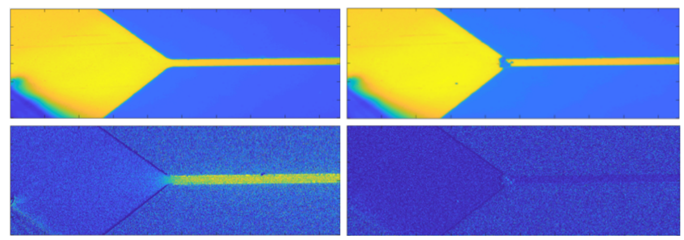

Fig. 1: Sample pairs of optical (top) and thermal (bottom) images for an aluminium device during EM. The first column shows the image pairs at the beginning of the current flow, and the second column shows the image pairs after device failure. EM creates a hole in the metal (visible in the optical image on top-right) that grows over time and eventually creates a failure. Our goal is to predict the failure location and time of failure from a single image before the failure occurs.

## I. Introduction

The modern world relies on the correct operation of innumerable electronic systems that are the backbone of everything from consumer devices to critical government and military infrastructure. The reliability of these systems, in turn, depends on the dependability of countless integrated circuits and other microelectronic devices. Detecting potential failures in these devices before they happen, and predicting their remaining lifespan before failure, is thus of critical importance.

Electromigration (EM), the process of displacing atoms in metals due to current flow [1], [2], causes interconnect failures in electronic circuits and is a major reliability concern as devices become smaller and electrical current densities in chip interconnects continue to rise. This EM phenomenon causes the metal to transport in the direction of charge-carrier (electrons) flow. This leads to migration of the metallic ions from the cathode (negative terminal) towards the anode (positive) terminal. Voids start to form near the cathode, eventually leading to an open circuit. While ion accumulations (hillocks) happen near the anode, which may lead to shorts with neighbouring interconnects. An example of the voids can be seen in Figure 1 (most apparent in top-right image). While EM has been studied for decades [1], [3], [4], [5], it is a complex process to model or test: finite element modeling of EM turns out to be so computationally expensive that it is practically infeasible.

In this paper, we explore the alternative approach of using Deep Neural Networks (DNNs) to model EM, which could offer significantly reduced test times, permit proactive hardware replacements, and potentially yield important insights into the physics of the EM process. While our long-term goal is to predict failures in arbitrary microelectronic devices, here we take the first step by considering aluminium ($Al$) interconnects that fail after a certain number of hours of continuous current flow. We systemically capture multimodal imagery (optical and thermal) during the entire duration of the flow until failure, creating a dataset of images of the device with paired known remaining lifespan as ground truth. Then, we propose two deep neural network models for solving the failure location identification and failure time prediction problems. We formulate the problem of finding the failure location of a device as an image segmentation problem, which we solve using a convolutional neural network (CNN). Finally, we cast the device age prediction task as a regression problem and solve it using a combination of convolutional neural network (CNN) and recurrent neural network (RNN).

To the best of our knowledge, our work is the first that pursues a machine learning-based approach for predicting EM failures from experimental optical and thermal images. EM is a complex process to model or test. We exploited two existing powerful paradigms, convolutional neural network and recurrent neural network, and showed how they can learn from optical and thermal images to model failure location and lifespan.

## II. Related Work

Object detection and segmentation have been extensively studied in computer vision, and have particularly benefited from the rise of deep neural networks. Object detection identifies objects and their locations, typically in terms of bounding boxes. Popular deep learning-based object detection methods include SSD [6], YOLO [7], and Mask R-CNN [8].

In contrast to detection, the goal of semantic segmentation is to precisely localize regions of interest with detailed pixel-wise annotations. FCN [9] is one of the most popular end-to-end trainable convolutional neural network for semantic segmentation, and inspired other methods such as SegNet [10]. U-Net, proposed for segmenting medical images, introduced several architectural innovations [11]. Others were inspired by U-Net's encoder-decoder architecture, including UNet++ [12] and 3D-DenseUNet [13]. 3D-DenseNet embedded a recurrent module – Convolutional LSTMs – at the bottleneck layer to capture the inter-slice continuity of a 3D volume. A combination of the convolutional and recurrent modules was also found beneficial in other tasks. For autonomous driving, Sen et al. [14] proposed a CNN + LSTM-based model to predict the angle of the steering wheel from images. De Fauw et al. [15] learned a tissue segmentation method for diagnosing retinal diseases, matching or even exceeding the performance of medical experts.

Deep learning methods are also being used in sensing modalities other than visible spectrum, including X-ray, MRI (Magnetic Resonance Imaging), and Lidar (Light Detection And Ranging) [16], [17], [18]. Wang et al. [16] classified X-ray scattering images with CNN and Convolutional autoencoders. Although visually very different from the visible spectrum, these X-ray scattering images are useful for material discovery and analysis. Guan et el. [19] introduced a CNN model for automatically annotating large-scale X-ray scattering images. Luo et al. [20] proposed a hybrid spatial and temporal neural network architecture to detect defects in infrared thermal images.

More directly related to our work, deep learning-based techniques have been adopted for EM analysis. Lin et al. [21] used neural networks to analyze the interconnect reliability of complicated integrated circuits (ICs) by modeling power amplifier circuits. The authors argued that the use of neural networks combined with existing reliability simulations significantly sped up the analysis. Their neural network model could help designers find the optimal transistor size and working conditions given a specific atomic flux divergence – a fundamental factor causing EM – while the sensitivity analysis can identify sensitive variables which can improve reliability. Kim et al. [22] proposed a novel cross-layer approach to optimizing the energy of a data center subject to long-term reliability and performance constraints. They proposed a novel physics-based EM model for a more accurate EM assessment of power grid networks at the chip level. An adaptive Q-learning-based reinforcement learning method was proposed in order to optimize the energy and reliability of the data center.

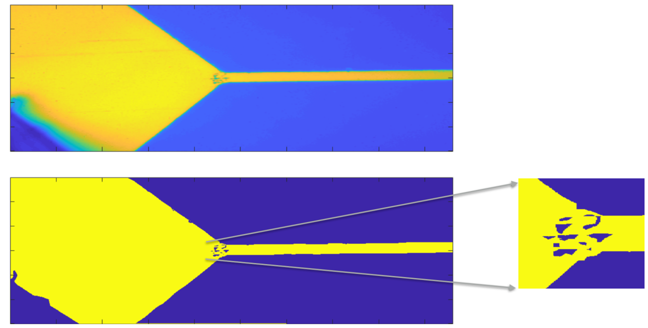

Fig. 2: Optical image (top) and corresponding binary segmentation map (bottom). A magnification of the segmentation (right) shows fragmentation caused by EM.

## III. Method

We propose a deep neural network-based model for finding the location of a device failure and predicting the remaining lifespan from images of the microelectronic device. A simple solution would be designing a single network that shares early layers for feature extraction but then separates into two task-specific branches. However, training such a network jointly in a supervised setting would require dense annotations — a single number for lifespan but pixel-by-pixel labels for segmentation — for each training image in both tasks. While device age or lifespan is straightforward to obtain from controlled lab experiments, it is very tedious and expensive to obtain pixel-level annotations for all training images.

Instead of jointly solving both tasks together, we thus propose separate deep neural network-based models for the two problems. Both networks are built on the top of a common convolutional neural network backbone. We also observed that the two problems require different amounts of supervision: the regression problem – the age of the device – requires more image-level annotations (which are easy to obtain), while solving the segmentation problem – finding the failure location – requires fewer images, but they must be annotated at the pixel-level (each image has around 100,000 pixels and labeling a few images results in sufficient labels to update the network parameters). We now discuss the two components of our approach, the Segmentation Network and the Age Prediction Network.

### A. Segmentation Network

We pose our failure localization problem as a binary image segmentation problem, in which we separate pixels into foreground (conductive metal) and background, and solved it with a convolutional neural network. We used the Optical (visual spectrum) images (Figure 2) to train our segmentation network. We defined foreground to be the region of current-conductive material (aluminium), as shown in the top image

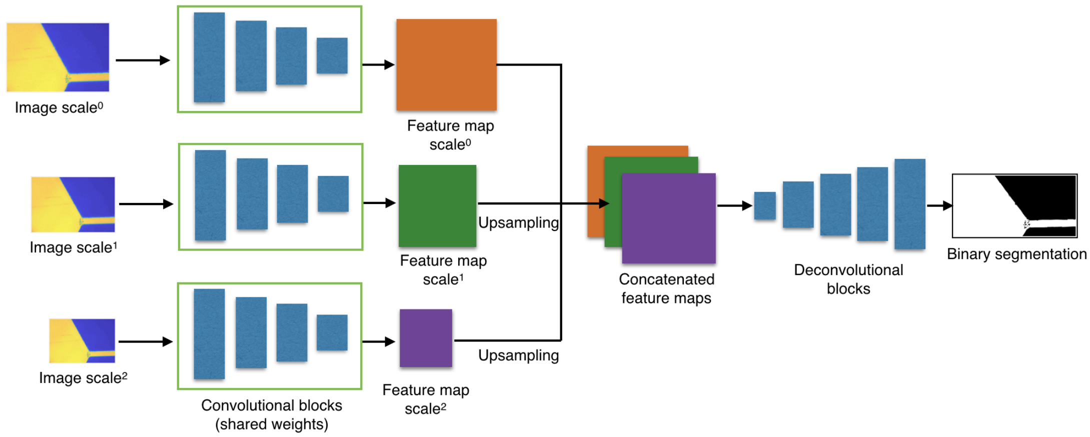


Fig. 3: Our segmentation network with multiscale encoder-decoder architecture.

of Figure 2 (the middle region in yellow color). The remaining region is denoted as background. We manually annotated some optical images at the pixel-level to prepare our ground truth, using the 'Object Labeling Tool' from Hoiem [23]. Figure 2 shows a sample annotated image, where darker pixels are background and lighter pixels are foreground. Notice the fragmentation due to EM that occurs in the bottleneck: we want to identify these voids because they will eventually lead to device failure.

We followed the encoder-decoder architecture of Lim et al. [24] to segment the image into foreground and background pixels. An image is fed through a series of convolutional layers at three different scales. The response maps are then concatenated together and transferred through another series of transposed-convolutional layers to recover the segmentation map. Figure 3 demonstrates an overview of the network architecture. The details of the two modules along with the network parameters are described in Table I. These convolutional blocks are also part of our device age prediction network (described below), so the two networks share common convolutional blocks that are beneficial for both tasks. We generate three different scales of the input image using a Gaussian pyramid with a sigma factor of 2. These images at multiple scales are passed through the convolutional modules. The weights of the convolutional blocks are shared among the inputs. The first three convolutional blocks are initialized with the pre-trained weights of VGG16 [25] and their weights are not updated during backpropagation. The weights of the remaining convolutional blocks along with the deconvolutional layers are learned from scratch.

| Layer | Description |
|---|---|
| Conv Block 1 | Conv(64)-Conv(64) -Maxpool(64) |
| Conv Block 2 | Conv(128)-Conv(128)-Maxpool(64) |
| Conv Block 3 | Conv(256)-Conv(256)-Maxpool(256) |
| Conv Block 4 | Conv(512)-Conv(512)-Conv(512) |
| Deconv Block 1 | TrConv(64)-TrConv(64)-TrConv(512) |
| Deconv Block 2 | TrConv(64)-TrConv(64)-TrConv(256) |
| Deconv Block 3 | TrConv(64)-TrConv(64)-TrConv(128) |
| Deconv Block 4 | TrConv(64) |
| Deconv Block 5 | TrConv(1) |

TABLE I: Our encoder-decoder style segmentation network architecture consisting of the convolutional blocks (encoder) followed by deconvolutional blocks. We used $ReLU$ nonlinearities after each convolution or transposed-convolution layer and $Dropout$ layers between the convolutional layers in Conv Block 4.

### B. Device Age Prediction Network

Our next goal is to predict the age (percentage of elapsed lifespan) given an image of a device. We formulate this as a regression problem, and our network architecture consists of a combination of a convolutional neural network (CNN) to extract spatial features and a recurrent neural network (RNN) to model the temporal dependencies between the frames. The architecture is shown in Figure 4. The convolutional network consists of the subset of convolutional blocks (first four blocks) that were used in our segmentation network. We added a fifth convolutional block with 32 convolutional layers followed by a ReLU nonlinearity layer. These convolutional blocks were followed by a series of fully connected layers, which we call our regression module. We appended our recurrent neural network within the regression module to encode our temporal features, and in particular gated recurrent units (GRUs) [26]. long short-term memory (LSTM) is a popular choice but we experimentally found that gated recurrent unit (GRU) performed better for the device age prediction task.

For a sequence of images $I_1, I_2, ..., I_T,$ assume the features from the convolutional module are denoted as $f_1, f_1, ..., f_T$. The sequence of output hidden states corresponding to these features is denoted as $h_1, h_1, ..., h_T$. For each $t \in T$, the GRU cell updates are given by,

$$\begin{aligned} z_t &= \sigma(W_z f_t + U_z h_{t-1}) \\ r_t &= \sigma(W_r f_t + U_r h_{t-1}) \\ h_t &= \tanh(W_h f_t + r_t \cdot U_h h_{t-1}) \\ h_t &= z_t \cdot h_t + (1 - z_t) \cdot h_t, \end{aligned} \tag{1}$$

where $W, U$ are learnable hidden layer parameters and $\cdot$ is element-wise matrix product. $z_t, r_t, h_t$, and $h_t$ are the update

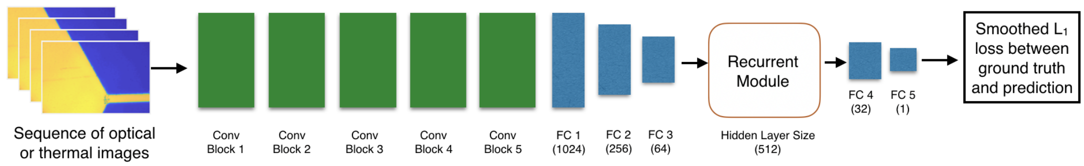


Fig. 4: Device age prediction network combining convolutional neural network (CNN) and recurrent neural network (RNN) to estimate the age of the aluminium $Al$ line.

gate, reset gate, new memory content gate, and hidden state, respectively.

## IV. Experiments

### A. Dataset

In the laboratory setting, we have captured a series of optical and thermal images of aluminum ($Al$) lines. In a typical experimental run, the $Al$ line is excited with a fixed electrical current that generates heat due to Joule heating. The wire undergoes EM and eventually breaks. Thermal and optical image pairs are saved every minute until the device failure, thus systematically recording the time evolution of temperature fields and defect generation around the failure region. The images were captured using thermoreflectance (TR) thermal imaging technique — a non-contact optical technique that generates high-resolution thermal images (spatial temperature distributions of the specimen's top surface). The technique uses visible light to generate thermal images with $\sim$10x the spatial resolution of infra-red (IR) thermal imaging. In this specific experiment, a 100$\times$ (NA=0.80) air objective was used under $\lambda$=780 nm illumination, yielding $\sim$400 nm of spatial resolution. Thermal image resolution was $\sim 0.1\,\mathrm{K}$ (the setup is capable of achieving $1\,\mathrm{mK}$ resolution with longer time-averaging). The detailed experimental setting, and an overview of the TR technique, are described in [27]. Moreover, a detailed finite element model that highlights our $Al$ lines complex EM physics can be found in [28]. Though TR imaging provides the surface's temperature distribution, buried interconnects can also be studied as long as their heating can create a surface temperature rise of a fraction of a degree. Since thermoreflectance technique can characterize transient heating with $50\,\mathrm{ns}$ time resolution, it is possible to study localized heating at different depths. For example, defects in different metallization layers in an IC chip have been characterized as they show distinct transient heating profile [29].

Here, we study three aging experiments. In each one, the electrical current excitation for the $Al$ line was fixed, but the overall environment (ambient) temperature was varied. The ambient temperatures are 69°C, 55°C, and 33°C for the video sequences EM 1, EM 2, and EM 3 respectively (in Table II). The higher the ambient temperature, the quicker the device fails. Failure times ranged from 2.6 hours for video sequence EM 1 to 8.2 hours for video sequence EM 3 (in Table II). Each experiment has 133 to 420 pairs of optical and thermal images. We report the total number of image-pairs in Table II. Instead of predicting the failure time directly, we estimate the age of the $Al$ line, in terms of percentage of lifespan. For a frame $t \in T$, this percentage of age is $a_t = 100 * t/T$ where $a_t$ is the variable we want to estimate for the $t^{th}$ frame and $T$ is the frame number in which failure occurs, e.g., 133 for EM 1 video sequence.

| | # frames | | | | |
|---|---|---|---|---|---|
| **Video Sequence** | **Train** | **Val** | **Test** | **Total** | **Hours to failure** |
| EM 1 | 80 | 13 | 40 | 133 | 2.58 |
| EM 2 | 124 | 21 | 61 | 206 | 3.97 |
| EM 3 | 252 | 42 | 126 | 420 | 8.18 |

TABLE II: Statistics of video sequences in our dataset.

### B. Segmentation

We trained our segmentation network using the eight randomly-selected annotated images from the video sequence EM 1. The size of each image is 1401x701 pixels. We implemented our network in PyTorch, and trained it for 30 epochs using binary cross-entropy loss. RMSProp solver was used for the optimization of the network parameters with an initial learning rate of 0.0001 and a batch size of 1. The learning rate was decayed by a factor of 0.1 with every 6 epochs.

Segmentation was evaluated using the same trained model on all three video sequences. Since we have a two class labeling problem in our segmentation task, we evaluated using the standard metrics used for semantic segmentation: 1) *per-class accuracy*, which measures the fraction of correctly predicted pixels for each class, and 2) *intersection over union (IoU)*, which computes the ratio of the size of the intersection of ground truth label and estimated label regions to the size of the union between the ground truth and the estimated label. To obtain a quantitative measure of our segmentation performance, we additionally annotated 8, 9 and 8 test images from videos EM 1, EM 2, and EM 3 respectively.

We evaluated our segmentation using both metrics and report the numbers in Table III. Our segmentation can predict correctly the background region, which includes the 'void' regions due to EM – our region of interest. Figure 5 shows visualizations of the results on the EM 3 test samples. We notice that our approach can segment the input optical test image progressively, predicting more of the 'void' regions as the experiment reaches towards total failure. Our segmentation method tends to predict more accurate locations towards the end of the device's lifespan, whereas sparse tiny 'void' regions that occur long before the actual breakdown are ore difficult.

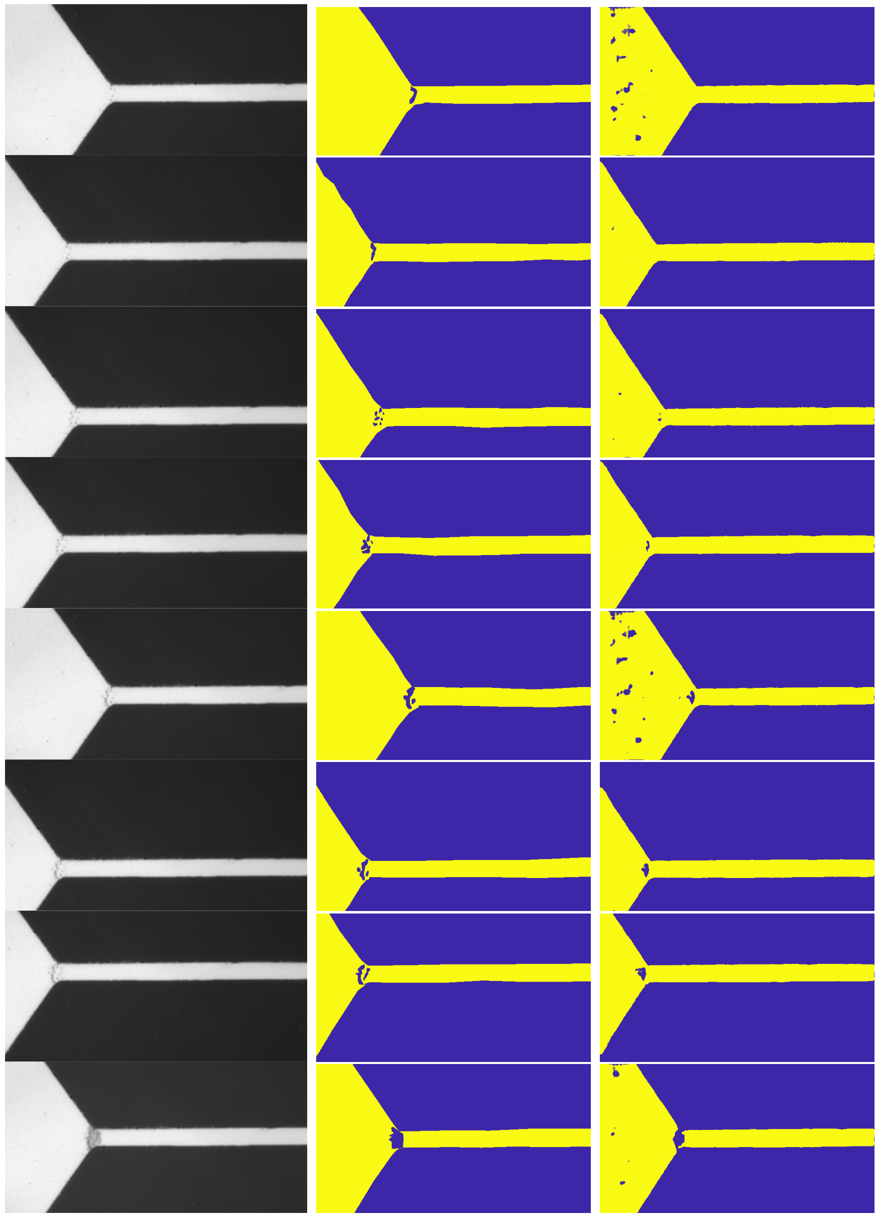

Fig. 5: Finding the device failure locations from our segmentation network (best viewed in color). Input optical (visual-spectrum) images (first column) are fed through our network which in turn predicts the binary foreground vs background segmentation (third column). The ground-truth maps are shown in the middle column. Each row shows a separate image in a video sequence (EM 3).

This problem can be partly attributed to the annotation process which can be improved to capture more precise boundaries on those sparse 'void' regions.

### C. Device Age Prediction Network

We trained our device age prediction network for each video sequence separately. Each video sequence consists of pairs of optical and thermal images. The image dimensions are 301x301 pixels. We partitioned the frames in a video sequence into train/val/test splits as shown in Table II. This regression network measures prediction accuracy in terms of two metrics: 1) Root Mean Square Error ($RMSE$), and 2) Mean Average Error ($MAE$). These two metrics are computed for a given video sequence from the individual frame's predicted and ground truth age,

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(Y_i^g - Y_i^p)^2} \tag{2}$$

$$MAE = \frac{1}{T}\sum_{i=1}^{T}|Y_i^g - Y_i^p|, \tag{3}$$

where $Y_i^g$ and $Y_i^p$ are the ground truth and predicted age of the $i^{th}$ frame and these values are computed for all the frames $i \in T$.

| | Per-class (↑) | | | IoU (↑) | | | |
|---|---|---|---|---|---|---|---|
| | BG | FG | Avg | BG | FG | Avg | Global Acc |
| EM 1 | 0.989 | 0.989 | 0.989 | 0.979 | 0.979 | 0.979 | 0.989 |
| EM 2 | 0.991 | 0.978 | 0.985 | 0.979 | 0.964 | 0.971 | 0.986 |
| EM 3 | 0.993 | 0.968 | 0.981 | 0.983 | 0.950 | 0.966 | 0.987 |

TABLE III: Evaluation of failure location prediction via our segmentation network.

| | | RMSE (↓) | | | | MAE (↓) | | | |
|---|---|---|---|---|---|---|---|---|---|
| Image Modality | Recurrent Module | EM 1 | EM 2 | EM 3 | Average | EM 1 | EM 2 | EM 3 | Average |
| Thermal | LSTM | 8.00 | 15.87 | 6.61 | 10.16 | 6.33 | 9.20 | 4.92 | 6.82 |
| | GRU | 11.09 | 17.86 | 6.74 | 11.90 | 8.59 | 8.53 | 5.39 | 7.50 |
| Optical | LSTM | 7.07 | 5.76 | **3.83** | 5.55 | 5.17 | 4.30 | **2.79** | 4.09 |
| | GRU | **5.83** | **5.14** | 3.98 | **4.98** | **4.62** | **4.28** | 2.99 | **3.96** |

TABLE IV: Results of our device age prediction network in RMSE and MAE metrics (lower is better). We analyzed the network with either optical or thermal images for all three video sequences separately. Our model with the GRU recurrent module on optical images achieved the best performance on average in both metrics.

| | | EM 1 | | EM 2 | | EM 3 | |
|---|---|---|---|---|---|---|---|
| Method | Image Modality | **RMSE** (↓) | **MAE** (↓) | **RMSE** (↓) | **MAE** (↓) | **RMSE** (↓) | **MAE** (↓) |
| CNN baseline | Thermal | 13.59 | 8.95 | 12.85 | 8.82 | 12.43 | 10.86 |
| | Optical | 8.77 | 7.14 | 8.49 | 7.79 | 9.66 | 8.89 |
| Fusion baseline (concat) | Optical + Thermal | 21.87 | 19.25 | 20.78 | 18.97 | 16.27 | 14.81 |
| Fusion baseline (add) | Optical + Thermal | 10.5 | 7.99 | 12.0 | 10.58 | 12.35 | 11.25 |
| Our method (GRU) | Optical | **5.83** | **4.62** | **5.14** | **4.28** | **3.98** | **2.99** |

TABLE V: Comparison of our device age prediction network's performance in RMSE and MAE metrics against two other baselines. We separately trained the CNN baseline with optical or thermal images for all three video sequences separately. Our fusion baseline learns jointly from both modalities where the weights of network parameters are randomly initialized. Our model with the GRU recurrent module on optical images performs better than two baselines in all video sequences.

As mentioned in Section III-B, the Age Prediction Network contains a series of convolutional blocks followed by three fully connected layers. After this last fully connected layer, we insert our recurrent GRU module. We used a unidirectional GRU with one hidden layer of size 512. This GRU module is connected with another 32 dimensional fully connected layer with ReLU non-linearity. A dropout layer is also added after this layer. Finally, we add another single fully-connected node to predict the age of the device. The complete network is shown in Figure 4. We compute the smoothed $L_1$ loss between the predicted age and the ground truth age. The loss gradients are computed and backpropagated to update the weights of the network, and the Adam solver was used for the optimization with an initial learning rate of $10^{-5}$. We trained our network for a maximum of 4000 epochs.

*1) Finding the Right Predictive Modality:* In our first experiment, we explored the predictive power of optical images and thermal images separately. We trained our network for each video sequence separately on the optical training images and measured error metrics on the test partition using Equations 2 and 3. The errors are reported in the last two rows of Table IV. Similarly, we trained our network from the thermal images on the three video sequences and computed the error metrics as reported in the first two rows in Table IV. Our GRU-based failure prediction network achieved $RMSE$ values of 5.83, 5.14, and 3.98 on the three video sequences with optical images, which is far better than $RMSE$ values of 11.09, 17.86, and 6.74 for the thermal images. Overall, the optical images are more predictive than thermal images, with an average error of 4.98 in comparison to 11.9. Similar trends have been observed for the $MAE$ error metric. As expected, we see that training on the dataset with more images creates a more accurate model.

We were surprised that the approach applied to thermal imagery worked significantly worse than with the optical imagery, despite the fact that heat is what causes the circuit failure. One possible reason is that the thermal images do not accurately reflect the true temperature; we normalize the thermal images to scale the pixel values to the range 0.0-1.0 using device-specific minimum and maximum temperature values. Another possible reason is that due to the device physics, only the temperature of the aluminum metal can be measured accurately, while the measured temperatures of the micro-voids forming in a stressed aluminium line are often not accurate.

We also experimented with replacing our recurrent module with a long short-term memory (LSTM) recurrent unit and trained the modified network. We maintained a similar number of parameters as used in our GRU recurrent module. The errors are reported in rows 1 and 3 for thermal images and optical im-

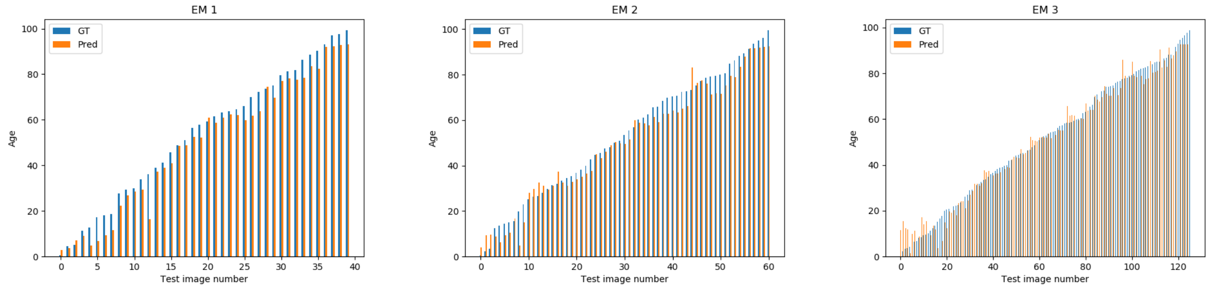


Fig. 6: Individual predictions from our GRU-based model on the test images for three video sequences. The ground truth ages are plotted in blue and predictions are in red (best viewed in color).

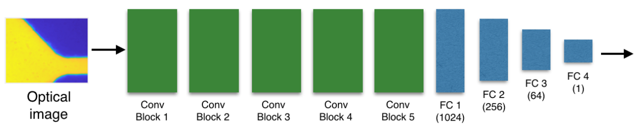


Fig. 7: The network architecture of our CNN baseline for failure prediction.

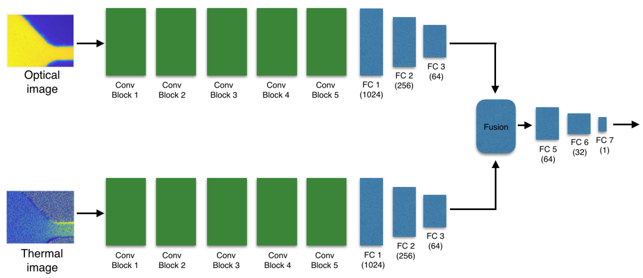


Fig. 8: The network architecture of our fusion baseline for failure prediction.

ages, respectively, in Table IV. For optical images, our GRU-based model reported an average error of 4.98 in comparison to 5.55 for the LSTM model in the $RMSE$ metric. A similar trend has been observed for $MAE$ metric, confirming that the GRU is the better choice. We also visualized individual predictions on the test partition for all three video sequences in Figure 6.

*2) Comparison with Other Baselines:* Now we turn our attention to comparing the performance of our device age prediction network with other baselines. To the best of our knowledge, ours is first in the domain of DNN-based EM failure analysis, so we devised other CNN-based baselines to compare the performance. We use two baseline architectures:

*CNN Baseline:* Our convolutional neural network (CNN) baseline is a model where we used only five convolutional blocks along with four fully connected layers sequentially. We train the network with the two image modalities, i.e., optical and thermal, separately using the smoothed $L_1$ loss function. The network architecture is depicted in Figure 7.

*Fusion Baseline:* Built on top of our CNN-based architecture, we designed a fusion of two CNN subnetworks, where the features are fused at a later stage benefiting the learning from both image modalities jointly during training. The network architecture is shown in Figure 8 and consists of two separate CNN-based subnetworks. The top branch receives optical images, while the bottom branch takes thermal images. We fuse the fully connected layers with a fusion operation. We experimented with different fusion operations between two feature vectors such as ‘add’, ‘concat’, etc. The fusion feature is fed through another series of fully-connected layers to predict the failure time. We train the network with the same loss function as used in our CNN baseline.

*Discussion:* The results of our baselines are reported in Table V. We consistently found that optical images are more useful in device age prediction, as our CNN baseline yields less RMSE errors with optical images obtaining 8.77, 8.49, and 9.66 for EM1, EM2, and EM3 video sequence, respectively. The RMSE errors are beyond 10 in all three experiments for thermal images. Our fusion baseline was designed in the hope of obtaining complementary signals from the two modalities. During experimentation, it did not result in improved performance compared to our CNN baseline that uses a single modality. The RMSE errors on the three video sequences are 10.5, 12, and 12.35 respectively, which is more than what our CNN baseline could predict. Moreover, we also experimented with different operations to find the right feature fusion. Our experiments revealed that the ‘add’ operation was the best choice. In contrast, our model could reduce the RMSE errors significantly in all three video sequences, as shown in Table V (last row).

## V. Conclusion

In this work, we demonstrated a deep neural network-based approach for understanding the EM phenomenon which is a fatal cause for interconnect failures in electronic circuits. Our approach provides an image-based machine learnable model for gaining important insights into the physics of the EM process which is otherwise a complex process to model. We developed an encoder-decoder style CNN model to accurately predict the failure location of a device. We also developed a recurrent neural network model to predict the age and remaining lifespan of the device. From our findings, we concluded that optical images are more useful in finding the failure location as well as predicting the age of the device during EM.

**Acknowledgments.** The work in this paper was funded by the Indiana Innovation Institute (IN3).

## References


[1] J. Black, “Physics of electromigration,” in *12th International Reliability Physics Symposium*, 1974.

[2] I. Blech and C. Herring, “Stress generation by electromigration,” in *Applied Physics Letters*, 1976.

[3] D. Pierce and P. Brusius, “Electromigration: A review,” in *Microelectronics Reliability*, 1997.

[4] K. Tu, “Recent advances on electromigration in very-large-scale-integration of interconnects,” in *Journal of Applied Physics*, 2003.

[5] F. dHeurle, “Electromigration and failure in electronics: An introduction,” in *Proceedings of IEEE*, 1971.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, “SSD: Single shot multibox detector,” in *European Conference on Computer Vision*, 2016.

[7] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” *arXiv*, 2018.

[8] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in *IEEE International Conference on Computer Vision*, 2017.

[9] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[10] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: a deep convolutional encloder-decoder architecture for scene segmentation,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[11] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

[12] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A nested U-Net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis*, 2018.

[13] U. Kamal, A. Rafi, R. Hoque, and M. Hasan, “Lung cancer tumor region segmentation using recurrent 3d-denseunet,” in *arXiv*, 2018.

[14] H. Sen, Y. Mi, D. Kangin, and N. Pugeault, “Aggregated sparse attention for steering angle prediction,” in *Proceedings of the International Conference on Pattern Recognition (ICPR)*. IEEE, 2018.

[15] “Clinically applicable deep learning for diagnosis and referral in retinal disease.” in *Journal of Nature Medicine*, 2018.

[16] B. Wang, K. Yager, D. Yu, and M. Hoai, “X-ray scattering image classification using deep learning,” in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2017.

[17] J. Qiu, Z. Cui, Y. Zhang, X. Zhang, S. Liu, B. Zeng, and M. Pollefeys, “DeepLiDAR: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[18] Z. Akkus, Galimzianova, A. A., Hoogi, D. Rubin, and B. Erickson, “Deep learning for brain mri segmentation: State of the art and future directions,” in *Journal of Digital Imaging*, 2017.

[19] Z. Guan, H. Qin, K. G. Yager, Y. Choo, and D. Yu, “Automatic x-ray scattering image annotation via double-view fourier-bessel convolutional networks,” in *British Machine Vision Conference 2018, BMVC*, 2018.

[20] Q. Luo, B. Gao, W. Woo, and Y. Yang, “Temporal and spatial deep learning network for infrared thermal defect detection,” in *NDT & E International*, 2019.

[21] Q. Lin, H. Fu, F. He, and Q. Cheng, “Interconnect reliability analysis for power amplifier based on artificial neural networks,” in *Journal of electronic testing*, 2016.

[22] T. Kim, Z. Sun, C. Cook, H. Zhao, R. Li, D. Wong, and S. Tan, “Cross-layer modeling and optimization for electromigration induced reliability,” in *In Proceedings of the 53rd Annual Design Automation Conference (DAC)*, 2016.

[23] D. Hoiem. Object annotation tool. [Online]. Available: http://dhoiem.cs.illinois.edu/software/objectLabelingTool.zip

[24] L. A. Lim and H. Y. Keles, “Foreground segmentation using convolutional neural networks for multiscale feature encoding,” in *Pattern Recognition Letters*, 2018.

[25] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *International Conference on Learning Representation (ICLR)*, 2015.

[26] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Gated feed-back recurrent neural networks.” in *International Conference on Machine Learning (ICML)*, 2015.

[27] S. Alajlouni, K. Maize, P. Bermel, and A. Shakouri, “Thermoreflectance Imaging of Electromigration in Aluminum Interconnects at Different Ambient Temperatures,” in *American Society of Mechanical Engineers Digital Collection*, Dec. 2019.

[28] Z. Wang, S. Alajlouni, P. Bermel, and A. Shakouri, “Explaining an unusual electromigration behavior—A comprehensive experimental and theoretical analysis using finite element method,” *Journal of Applied Physics*, vol. 129, no. 21, p. 214502, Jun. 2021. [Online]. Available: https://aip.scitation.org/doi/10.1063/5.0040526

[29] K. Yazawa, D. Kendig, and A. Shakouri, “Time-Resolved Thermoreflectance Imaging for Thermal Testing and Analysis,” San Jose, California, USA, Nov. 2013, pp. 194–202. [Online]. Available: https://dl.asminternational.org/istfa/proceedings/ISTFA2013/80224/194/11780