Vmin Prediction

Silicon Lifecycle Management (SLM) Use Case

Challenges

Choosing the right operating voltage for various digital blocks within a semiconductor device is one of the most important tasks faced by chip designers. The operating voltage has a major impact on performance, power consumption, and reliability. Increasing the voltage generally increases performance, but at the cost of more power, higher lifetime operating costs and shorter lifetime expectancy. The key challenge is selecting the right value for Vmin, which is defined as the minimum supply voltage at which each chip or individual block within a device can operate properly. Designers need to successfully identify the minimal voltage level that still enables safe device operation given the additional factors of die variation, environmental stress, aging, as well as different silicon-to-package characteristics. True minimum voltage also varies from device to device.

A common approach today for Vmin is to categorically add excessive guard-banding across all parts and as a result, this leads to less-than-optimal Vmin for at least a portion of the population. Conversely, performing exhaustive testing during System-Level Test (SLT) for identifying the true Vmin per part is cost prohibitive due to excessive time on the tester. Alternately, an attempt to create an optimal predictive model does not usually yield good results due to the lack and quality of silicon data and robustness of the Machine Learning (ML) models developed.

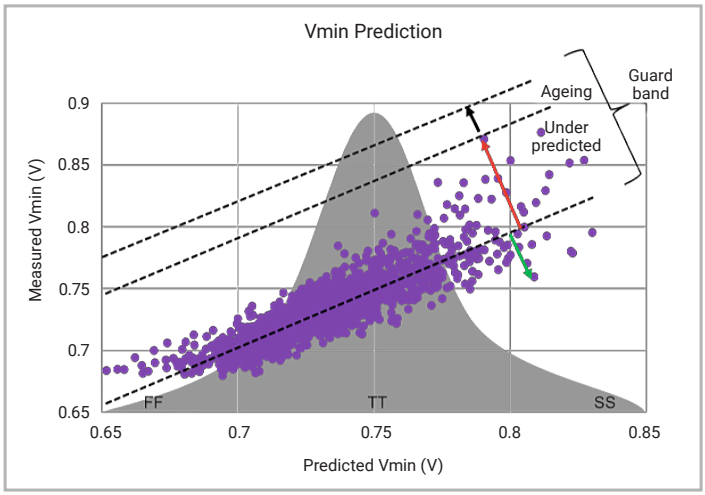

Figure 1. Existing Vmin prediction approach

Synopsys Solution

Vmin prediction is one of the key applications of the SLM Silicon.da Monitor Analytics solution and its fundamental goal is to improve the correlation between a predicted Vmin that comes from a model versus the measured Vmin. This enables the device to operate at lower power while still meeting the required performance by setting the device specific Vmin during testing on the test floor at the value of the predicted model.

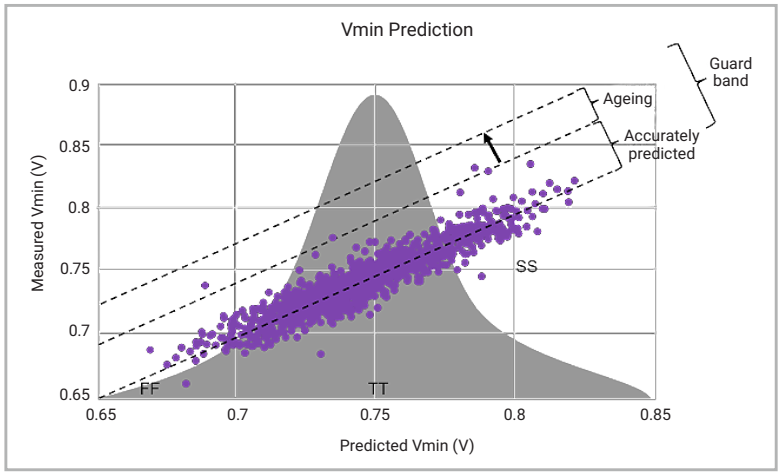

Figure 2. Synopsys Vmin prediction model

Figure 2. shows the improvement in accuracy of the Synopsys Vmin prediction model compared to existing models. The graph shows a tighter correlation between what is measured and what is predicted compared to previous approaches. This provides the following benefits to design teams:

- Identifies lowest voltage a device can run at while still meeting performance requirements
- Reduces test cost and minimize overall test time
- Improves overall device reliability

How it Works

A prediction model needs to be created, trained and verified by using specific datasets from several hundred actual chip samples during chip bring-up. The datasets combine data from embedded ring oscillators (ROs) collected during any or all of the following test stages: wafer sort (WS), final test (FT) and system-level test (SLT). These ROs can be the customer's proprietary IP or can come from a 3rd party or from Synopsys in the form of process detectors. Additionally, datasets from actual Vmin search measurement testing need to be collected during SLT and correlated with the RO datasets. It is also important that the data collected represent varied process results. For example, it could be beneficial to use a "corner lot" for the training phase of the prediction model.

Collectively, all of this data is then used together with ML to create, train and validate a Vmin prediction model yielding a tighter, more correlated optimized voltage per device. Now the product teams can use this predicted model during production of the next batch of new silicon by taking the RO measurements as input to the model and calculating the corresponding Vmin values to be used per device downstream in production and in field operation.

This Vmin method provides the best trade-off between accuracy and cost as described in more detail in the case study below.

Case Study—Procedural Steps for Vmin Prediction

This case study shows the procedural steps needed to deploy the Vmin Prediction model in a production environment.

The first step is to create and train the ML model on a predetermined set of devices. Note that actual measured Vmin testing needs to be done and the subsequent collected measured Vmin data used as a key contributor to train the model.

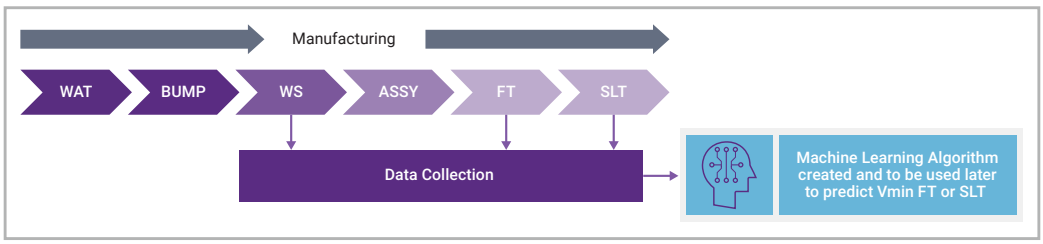

Figure 3. Generate the Machine Learning model offline

The second step is to apply the model on production data collected during wafer sort on the next batch of new silicon. In this case, the Vmin testing is no longer needed to be performed as the ML model will now be predicting the Vmin value to be used during FT and/or SLT. The predicted Vmin will be stored offline in memory until it is needed during FT or SLT.

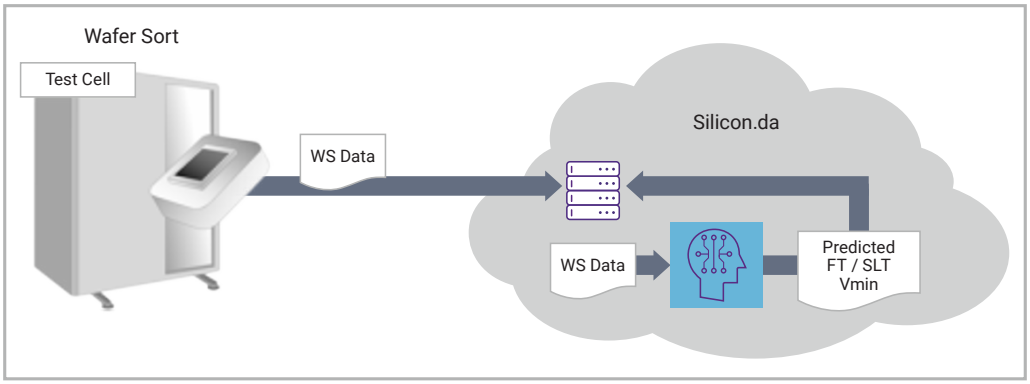

Figure 4. Use the Machine Learning model in wafer sort production

The third and final step utilizes the predicted Vmin during FT or SLT. The engineers managing the testing can either leave the predicted Vmin per device as the final Vmin to be used in the end product or the Vmin predicted values can be used as a starting point for further measured Vmin testing if they would like to improve upon their results across the full population of devices.

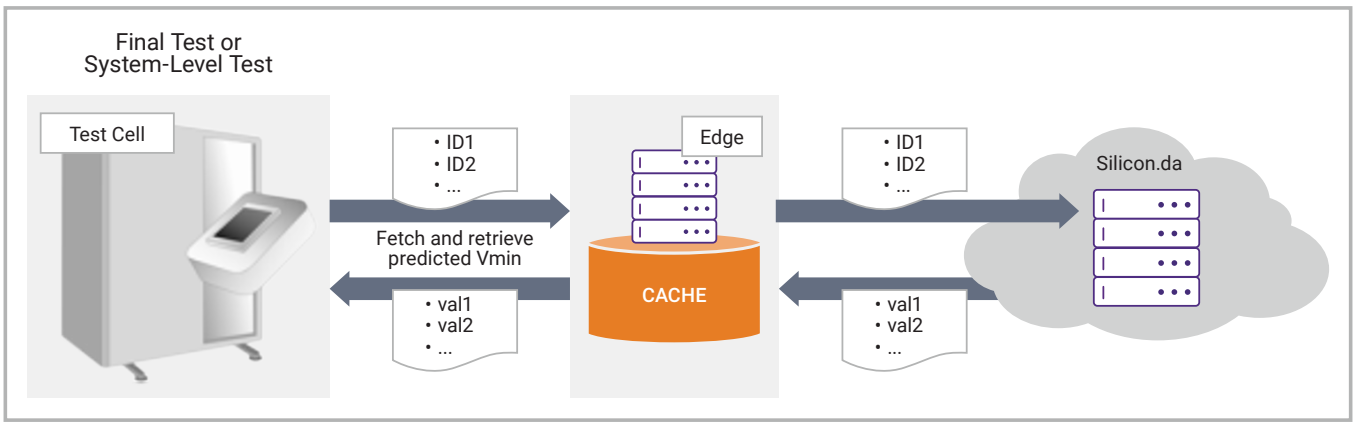

Figure 5. Use the predicted Vmin data in FT/SLT production