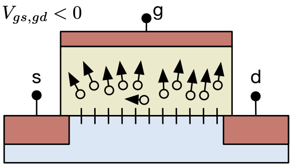
Analog IC Aging-induced Degradation Estimation via Heterogeneous Graph Convolutional Networks

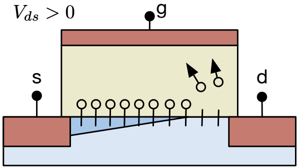
Tinghuan Chen[1], Qi Sun[1], Canhui Zhan[2], Changze Liu[2], Huatao Yu[2], Bei Yu[1]

[1]The Chinese University of Hong Kong
[2]Hisilicon Technologies Co.

Transistor aging: BTI and HCI



BTI mechanism

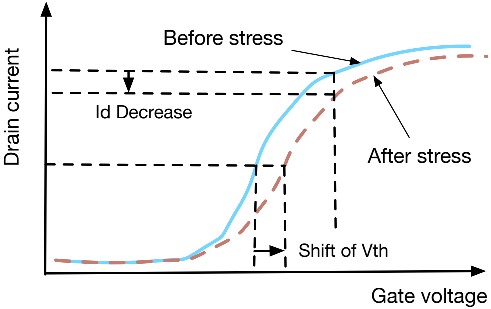


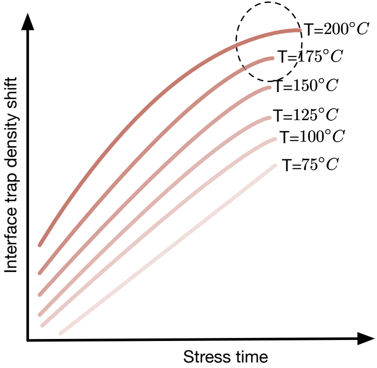
HCI mechanism

BTI and HCI mechanisms

- BTI mechanism is that accumulated holes in silicon/oxide interface result in breaking of Si-H bonds.
- HCI mechanism is that due to high electric field in drain side, hot carriers cause breaking of Si-H bond and traps oxide bulk.

Circuit performance degradation





Failure mechanism

- These aging effects cause transistor parameters, e.g., threshold voltage, to shift from their nominal values over time, resulting in gradual circuit failure.

Drawbacks of previous works

Analytical models

- A correct judgment on the reliability of the circuits heavily depends on the appropriate stress conditions for dynamic aging analysis. It is hard to find the appropriate stress conditions that lead to the actual degradation on the aging-prone transistors.
- While the static aging analysis causes inaccurate judgment on the aging-prone transistors since the dynamic stress conditions are completely ignored.
- It is time-consuming to achieve accurate detections when these models are implemented within the simulators.

Problem Formulation

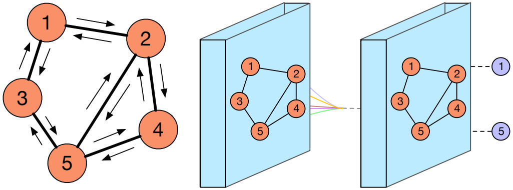
Definition 1

`dvtlin (HCI+BTI)` is defined as the shifting value of threshold voltage of the transistor from fresh to 10 years due to HCI and BTI.

Estimating `dvtlin` in Analog ICs

Given some analog IC post-layout netlists and a list containing all transistors with `dvtlin` obtained by the dynamic aging analysis as the training set, our task is training a model on the training set to fast and accurately estimate the `dvtlin` of transistors on the testing set while minimizing the estimation error.

Graph Convolutional Networks ∗



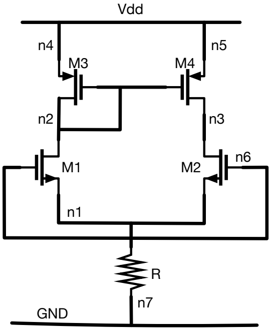
$$\mathbf{f}_i^{(l)} = \sigma\left(\text{CONCAT}(\mathbf{f}_i^{(l-1)}, \sum_{j\in\mathcal{N}(i)} \mathbf{f}_j^{(l-1)})\mathbf{W}^{(l)}\right)$$

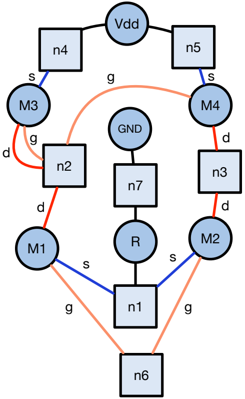
$$\mathbf{F}^{(l)} = \sigma\left(\text{CONCAT}(\mathbf{F}^{(l-1)}, \mathbf{A}\mathbf{F}^{(l-1)})\mathbf{W}^{(l)}\right)$$

$$\begin{bmatrix}\mathbf{f}_1^{(l)}\\ \mathbf{f}_2^{(l)}\\ \mathbf{f}_3^{(l)}\\ \mathbf{f}_4^{(l)}\\ \mathbf{f}_5^{(l)}\end{bmatrix} = \sigma\left(\text{CONCAT}\left(\begin{bmatrix}\mathbf{f}_1^{(l-1)}\\ \mathbf{f}_2^{(l-1)}\\ \mathbf{f}_3^{(l-1)}\\ \mathbf{f}_4^{(l-1)}\\ \mathbf{f}_5^{(l-1)}\end{bmatrix}, \begin{bmatrix}0&1&1&0&0\\ 1&0&0&1&1\\ 1&0&0&0&1\\ 0&1&0&0&1\\ 0&1&1&1&0\end{bmatrix}\begin{bmatrix}\mathbf{f}_1^{(l-1)}\\ \mathbf{f}_2^{(l-1)}\\ \mathbf{f}_3^{(l-1)}\\ \mathbf{f}_4^{(l-1)}\\ \mathbf{f}_5^{(l-1)}\end{bmatrix}\right)\mathbf{W}^{(l)}\right)$$

∗Kipf+, "Semi-supervised classification with graph convolutional networks," ICLR, 2016.

Circuit and Graph Representation†





Heterogeneity of analog IC

- Typical analog IC netlists have heterogeneity since they contain multi-typed basic devices and multi-typed connection ports.

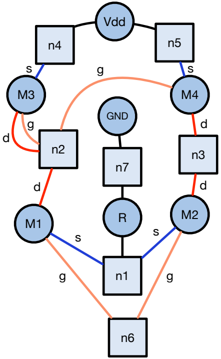
†Meissner+,“Feats: Framework for explorative analog topology synthesis,” IEEE TCAD, 2014.

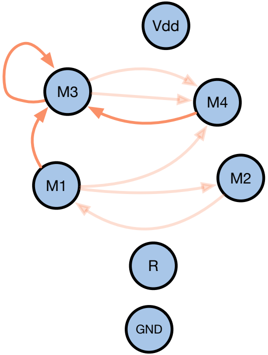
Heterogeneous directed multigraph

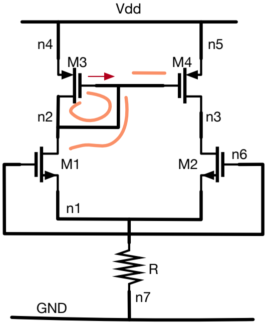
Heterogeneous directed multigraph

A heterogeneous directed multigraph is defined as a graph $\mathcal{HMG}_d(\mathcal{V}_{hmg}, \mathcal{E}_{hmg}, \mathcal{O}_{\mathcal{V}_{hmg}}, \mathcal{R}_{\mathcal{E}_{hmg}}, \varphi_{hmg})$, where $\mathcal{V}_{hmg}$ is the set of nodes, $\mathcal{E}_{hmg}$ is the multiset of edges. $\mathcal{O}_{\mathcal{V}_{hmg}}$ and $\mathcal{R}_{\mathcal{E}_{hmg}}$ represent the sets of node types and edge types, respectively. φ_{hmg} is adopted to assign each node in $\mathcal{V}_{hmg}$ to a node type, *i.e.*, $\varphi_{hmg}(\upsilon_i) \in \mathcal{O}_{\mathcal{V}_{hmg}}$ for $\forall \upsilon_i \in \mathcal{V}_{hmg}$. r indicates the edge type, such that $r \in \mathcal{R}_{\mathcal{E}_{hmg}}$. Each instance in $\mathcal{E}_{hmg}$ is $((\upsilon_i, \upsilon_j), r)$ and the ordered pair (υ_i, υ_j) satisfies $\upsilon_i, \upsilon_j \in \mathcal{V}_{hmg}$.

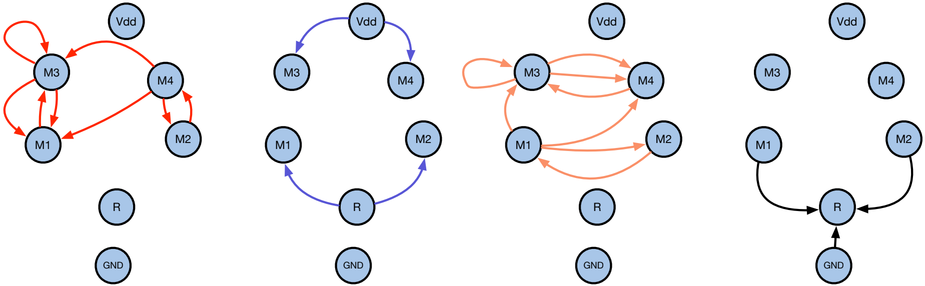
Heterogeneous Graph Representation







Heterogeneous Graph Representation



- One adjacency matrix is used to represent one type of edges in the heterogeneous multigraph.

Heterogeneous feature in analog IC

Heterogeneous feature

- Typical analog integrated circuit is composed of multiple types of basic devices, such as resistor, capacitor, inductor, transistor and diode.
- Each device has specific design parameters as attributes.
- A unified feature is used to encode all types of basic devices.
- It will miss some information with structural relations (edges) among multi-typed nodes as well as unstructured content associated with each node.

Latent space representation

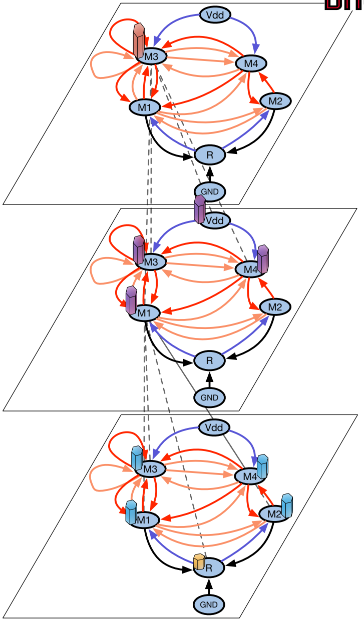
- A node $v_i \in \mathcal{V}_{hmg}$ whose node type is $t = \varphi_{hmg}(v_i) \in \mathcal{O}_{\mathcal{V}_{hmg}}$ and feature vector is $\mathbf{x}_{v_i} \in \mathbb{R}^{1\times h_t}$;
- A node-typed-related matrix $\mathbf{U}_t \in \mathbb{R}^{h_t \times \lambda}$ to map the feature vector with length h_t into a unified λ-dimension latent space, *i.e.*, $\mathbf{f}_{v_i}^{(0)} = \mathbf{x}_{v_i} \cdot \mathbf{U}_t \in \mathbb{R}^{1\times\lambda}$.

$$\mathbf{F}^{(0)} = \sum_{t \in \mathcal{O}_{\mathcal{V}_{hmg}}} \mathbf{X}_t \cdot \mathbf{U}_t \in \mathbb{R}^{|\mathcal{V}_{hmg}|\times\lambda}.$$

Embedding Generation

$$\mathbf{F}_{\mathcal{N}}^{(l-1)} = \left(\sum_{r\in\mathcal{R}_{\mathcal{E}_{hmg}}} w_r\mathbf{A}_r\right)\cdot\mathbf{F}^{(l-1)},$$

$$\mathbf{F}^{(l)} = \sigma\left(\mathrm{CONCAT}\left(\mathbf{F}_{\mathcal{N}}^{(l-1)},\mathbf{F}^{(l-1)}\right)\cdot\mathbf{W}^{(l)}\right).$$

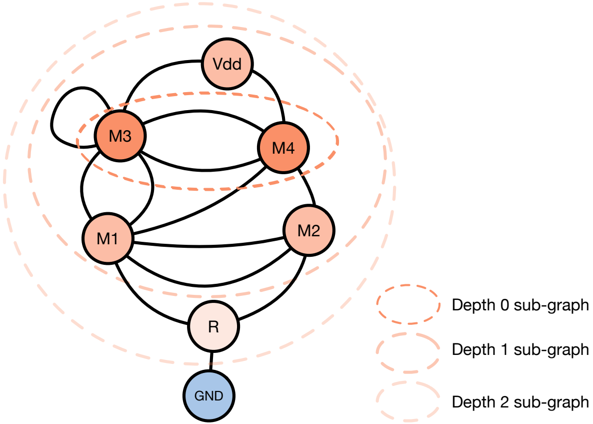


Node embedding

Embedding Generation of H-GCN

```
Require: $|\mathcal{R}_{\mathcal{E}_{hmg}}|$ adjacency matrices $\mathbf{A}_r$ and connection-type-related model coefficients $w_r$ with $\forall r \in \mathcal{R}_{\mathcal{E}_{hmg}}$; feature matrices of all nodes corresponding to each type of nodes $\mathbf{X}_t$ and node-type-related latent space mapping matrices $\mathbf{U}_t$ with $\forall t \in \mathcal{O}_{\mathcal{V}_{hmg}}$; Search depth $D$; non-linear activation function $\sigma(\cdot)$; Model coefficients matrices $\mathbf{W}^{(l)}$.
$\mathbf{F}^{(0)} \leftarrow \sum_{t \in \mathcal{O}_{\mathcal{V}_{hmg}}} \mathbf{X}_t \mathbf{U}_t$;
for $l = 0$ to $D - 1$ do
    $\mathbf{F}_{\mathcal{N}}^{(l)} \leftarrow \left( \sum_{r \in \mathcal{R}_{\mathcal{E}_{hmg}}} w_r \mathbf{A}_r \right) \cdot \mathbf{F}^{(l)}$;
    $\mathbf{F}^{(l+1)} \leftarrow \sigma \left( \mathrm{CONCAT} \left( \mathbf{F}_{\mathcal{N}}^{(l)}, \mathbf{F}^{(l)} \right) \cdot \mathbf{W}^{(l)} \right)$;
end for
return Embedding feature matrix $\mathbf{F}^{(D)}$.
```

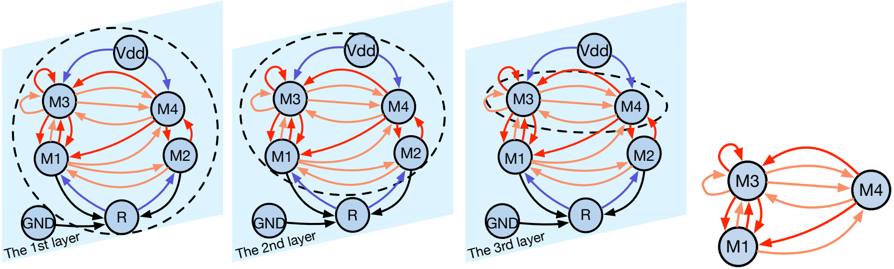
Large scale circuit netlist training



The Neighborhood Sampling Algorithm

```
Require: The input node set $\mathcal{V}_s$, neighborhood threshold $T$.
$\mathcal{V}^{(0)} \leftarrow \mathcal{V}_s$;
for $k = 1$ to $D$ do
    $\mathcal{V}^{(k)} \leftarrow \mathcal{V}^{(k-1)}$;
    for all $\mu \in \mathcal{V}^{(k-1)}$ do
        $\mathcal{V}^{(k)} \leftarrow \mathcal{V}^{(k)} \cup \mathcal{N}(\mu)$;
    end for
end for
if $|\mathcal{V}^{(D)}| > T$ then
    for all $v_i \in \mathcal{V}^{(D)} \backslash \mathcal{V}_s$ do
        Calculate $\mathcal{P}(v_i) = \frac{|\mathbf{A}(i,:)|+|\mathbf{A}(:,i)|-\mathbf{A}(i,i)}{|\mathbf{A}|}$;
    end for
    Sample $T - |\mathcal{V}^{(D)}|$ nodes according to the probability $\mathcal{P}$ to the set $\mathcal{V}_b$;
else
    $\mathcal{V}_b \leftarrow \mathcal{V}^{(D)}$;
end if
return The sampled node set $\mathcal{V}_b$;
```

The Neighborhood Sampling Algorithm



(a)

(b)

(a) Neighborhood expansions; (b) The sampled set $\mathcal{V}_b$.

Benchmark: Advanced industrial designs

Table: Statistics of Designs

Design	#trans.	#device	#net
1	4348	99009	18155
2	4382	99696	18299
3	3999	179758	31303
4	3998	185480	33819
5	523	31279	6002

Table: Types of Devices

Type	# Design parameter
MOS	51
MOS spice	75
DIO/ESD	8
Cap	12
R	6
VSource	1

- Industrial $5nm$ technology;
- PLL designs;
- Netlist files include Spectre and SPICE formats.

Baseline Methods

Method	Feature		Graph representation	
	Concat.	Latent	Homo.	Heter.
GPR	√		/	/
MLP	√		/	/
GCN	√		√	
Our H-GCN-concat	√			√
Our H-GCN		√		√

Implementation Details

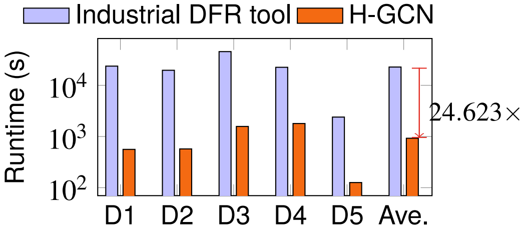
- Mean Square Error (MSE) function is used as the loss function;
- Weight decay hyperparameter: 10^{-7};
- Stochastic gradient descent for optimization;
- 2 H-GCN embedding layers + 3 fully-connected layers;
- Each output feature size: 512;
- The number of the training epoch: 500;
- Graph representation is implemented with Python and Networkx library;
- GCN models are implemented with TensorFlow.

Evaluation Metrics

- The aging performance is obtained from industrial DFR tool.
- Mean Absolute Error (absolute metric) and r2 score (relative metric) as metric to evaluate the performance.
- `dvtlin` is used to evaluate device aging performance.

Experimental Results

Design	GPR		MLP		GCN		Our H-GCN-concat		Our H-GCN	
	MAE	r^2 Score	MAE	r^2 Score	MAE	r^2 Score	MAE	r^2 Score	MAE	r^2 Score
1	3.394	0.366	3.037	0.370	1.416	0.618	1.307	0.775	**1.060**	**0.807**
2	3.481	0.448	3.283	0.540	1.395	0.601	1.355	0.763	**1.070**	**0.784**
3	5.041	0.257	4.736	0.415	4.404	0.558	2.771	0.760	**2.462**	**0.791**
4	4.985	0.245	4.850	0.401	4.408	0.548	2.811	0.749	**2.487**	**0.753**
5	4.641	0.393	4.230	0.595	2.344	0.792	1.861	0.818	**1.427**	**0.851**
Ave.	4.308	0.342	4.027	0.464	2.793	0.627	2.021	0.773	**1.701**	**0.797**



Conclusion

- We propose a H-GCN framework to detect aging prone devices.
- We propose a unified graph representation for analog IC netlist.
- The heterogeneous embedding generation is developed to map features of various devices to a unified feature space and multiple interconnections between the heterogeneous devices in an analog IC.
- Compared with traditional machine learning and graph learning methods, the proposed H-GCN can achieve more accurate estimations of aging-induced transistor degradation.
- Compared with an industrial dynamic aging DFR tool, our proposed H-GCN can achieve $24.623\times$ speedup on average.

Thank You