Parametric Modeling for Improved Die Area Estimation in the Life Cycle Assessment of Electronic Systems

Electronics Goes Green 2024+

Session B.3: Modelling of ICs and IC Packaging

Augustin WATTIEZ, Robin DETHIENNE, Anne-Sophie COLLIN, David BOL

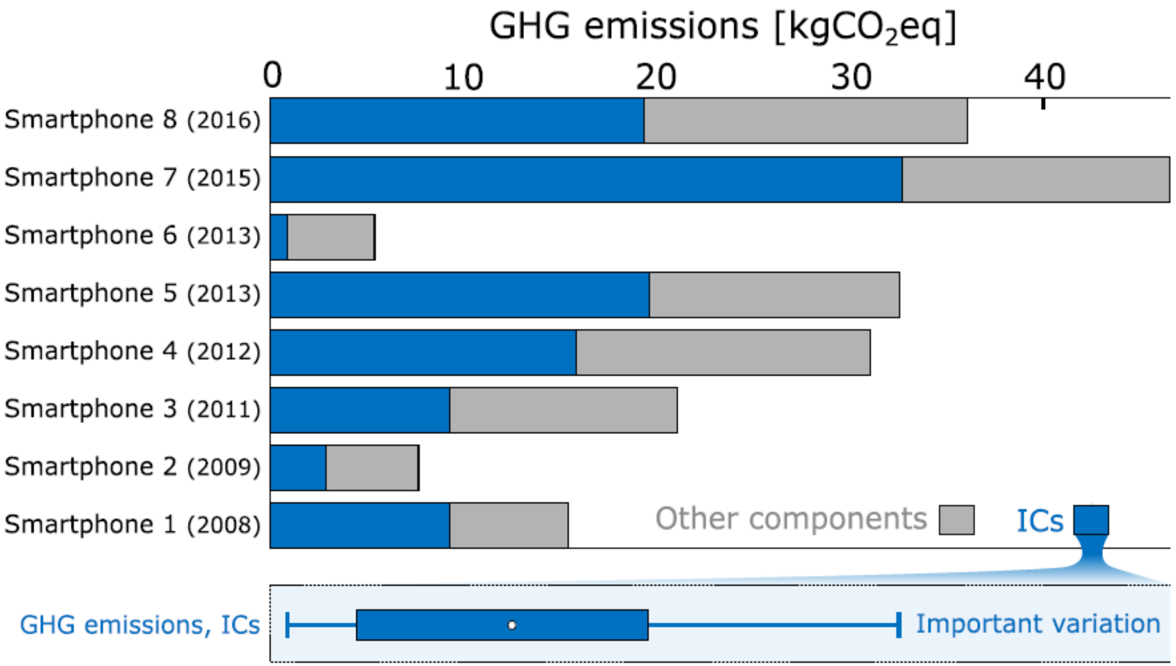
Augustin.wattiez@uclouvain.be

June 18, 2024

ELECTRONICS GOES GREEN 2024+
FROM SILICON TO SUSTAINABILITY
June 18 - 20, 2024 | Berlin, Germany

Motivation of this work

Modeling integrated circuits for a simple electronic system



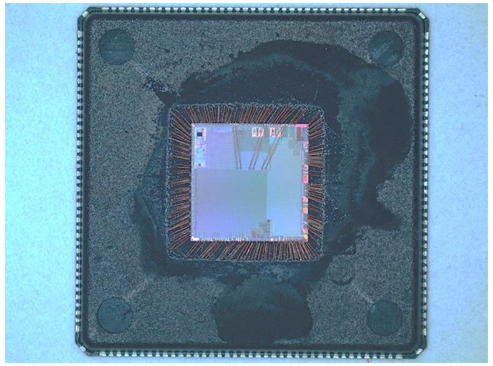
Integrated Circuits (ICs) often appear as **environmental hotspots** in the LCAs of **electronic systems**.
The **silicon die area** encapsulated in the package **dominates environmental impacts**.

Illustration adapted from: L-P. P.-V.P. Clément, Q. E. S. Jacquemotte, et L.M. Hilty, "Sources of variationin life cycle assessments of smartphones and tablet computers", Environ.Impact Assessment Rev., vol. 84, p. 106416, Sep. 2020, https://doi.org/10.1016/j.eiar.2020.106416.

Motivation of this work

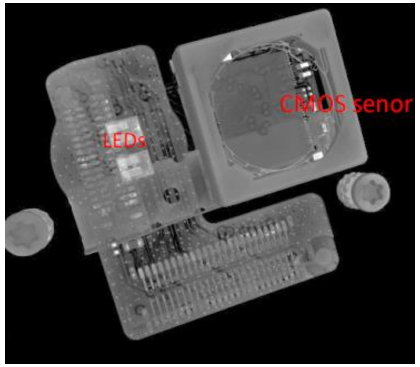
Modeling integrated circuits for a simple electronic system

Chemical Decapsulation



Mechanical Grinding

X-rays



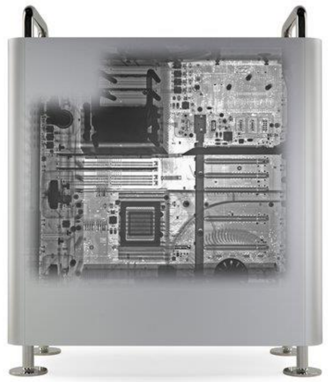
Method	Chemical decapsulation	Mechanical grinding	X-rays
Uncertainty on prediction	None	None	Low**
Modeling time/IC [min]	20 – 30*	30 – 40*	NA
Facility complexity	Moderate	Moderate	High
Destructive ?	YES	YES	NO

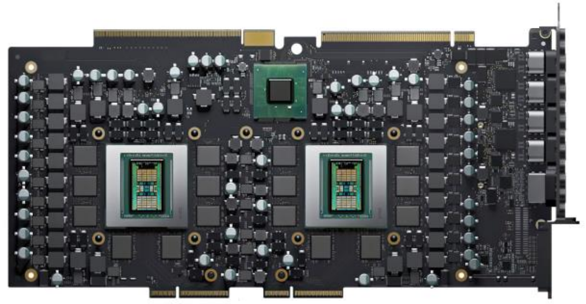
* Including circuit labeling and de-soldering
** It depends on package type

Outline

1. Modeling integrated circuits for complex electronic systems
2. Research questions and methodology
3. Parametric modeling
   I. Data visualization
   II. Linear univariate models
   III. Linear and non-linear multivariate models
   IV. Models comparison
4. Conclusions and future works
5. Taking a step back

1. Modeling integrated circuits for complex electronic systems

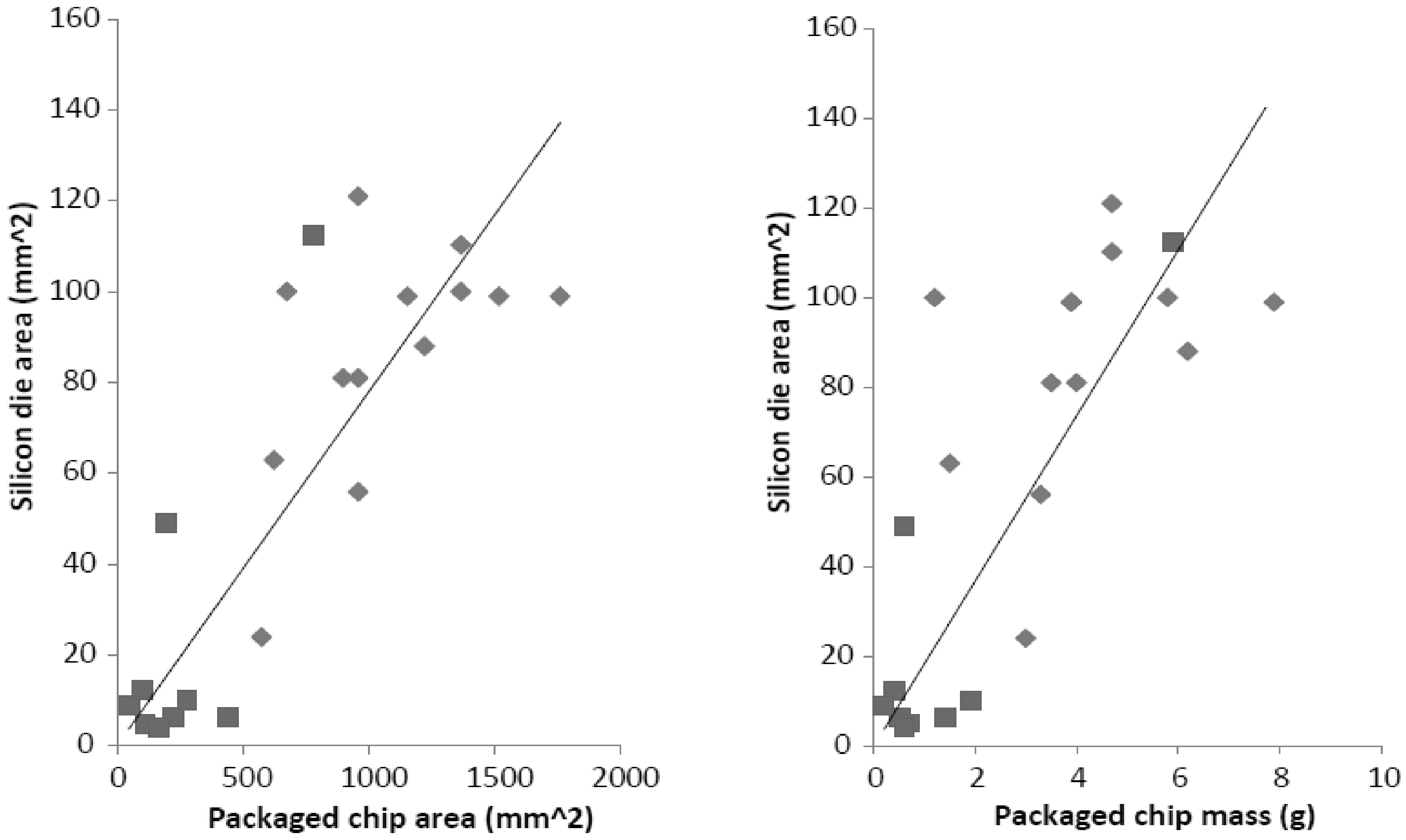




Mac Pro 2019. It typically contains **> 100 ICs.**

Method	Chemical decapsulation	Mechanical grinding	X-rays	Prediction models
Uncertainty on prediction	None (measurement)	None (measurement)	Low	**?**
Modeling time/IC [min]	20 – 30	30 – 40	?	Less than 1 min.
Modeling time – 100 ICs [working days]	**4.5 – 7**	**7 – 9.5**	?	**Less than 0.5**
Facility complexity	Moderate	Moderate	High	NA
Destructive ?	YES	YES	NO	NO

1. Modeling integrated circuits for complex electronic systems



Teehan, P. (2014). Integrative approaches to environmental life cycle assessment of consumer electronics and connected media (Doctoral dissertation, University of British Columbia).

2. Research questions and methodology

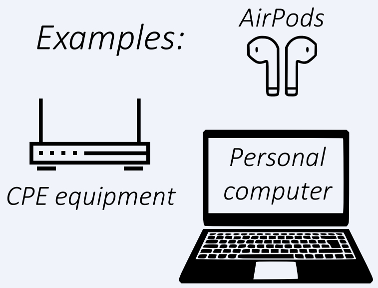
Two research questions:

1. Can the die area be better estimated based on an **extended feature set**?
2. Can **other models** be used to improve prediction?

Four-step methodology:

Identifying and **collecting** ICs in electronic systems

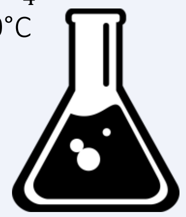
Examples:



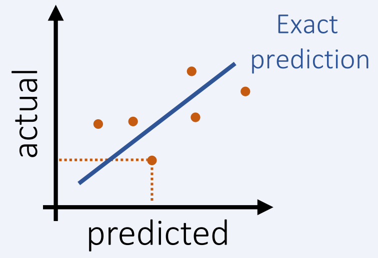
For **each IC**, **extracting** the following **features**:
- Package mass*
- Package area*
- Pin count
- Package type (class)
- Function (class)

Extracting **die area** through **chemical decapsulation**

H_2SO_4
170°C



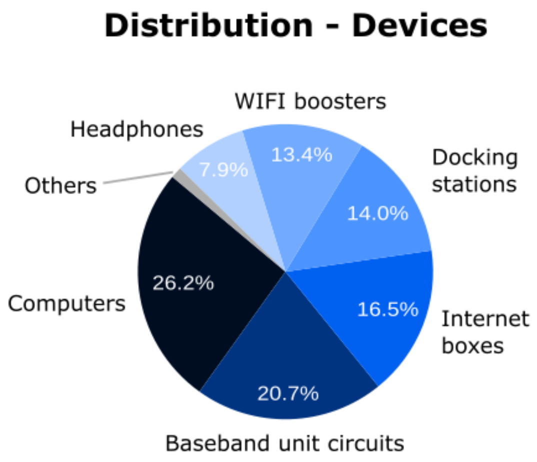
Regression between the **features** and the **die area**

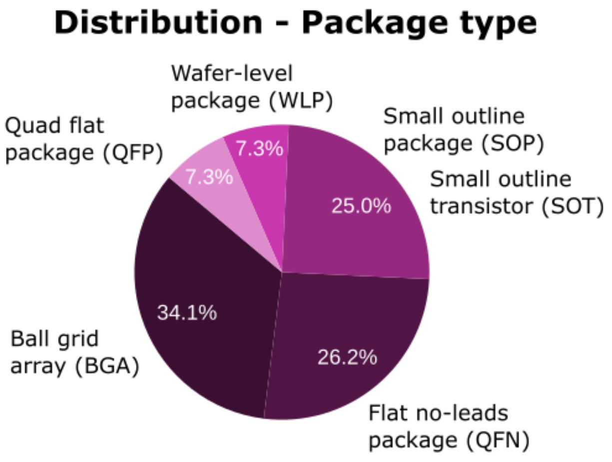


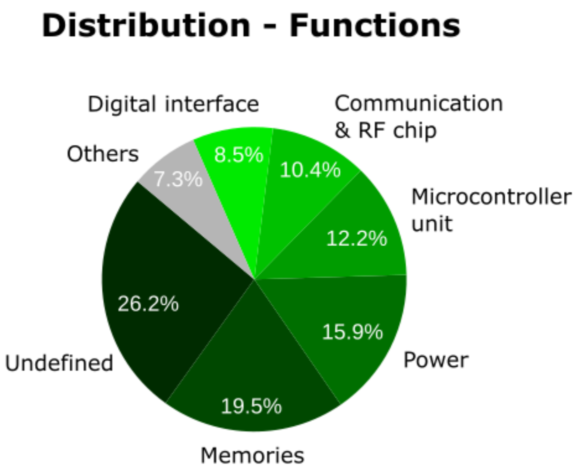
* Existing research only focusses on these parameters

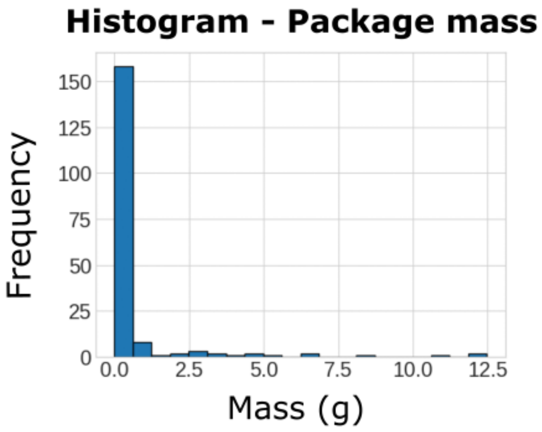
3. Parametric modeling

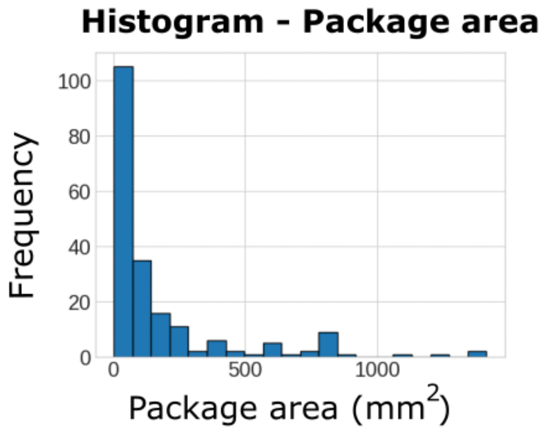
I. Data visualization – 200 ICs

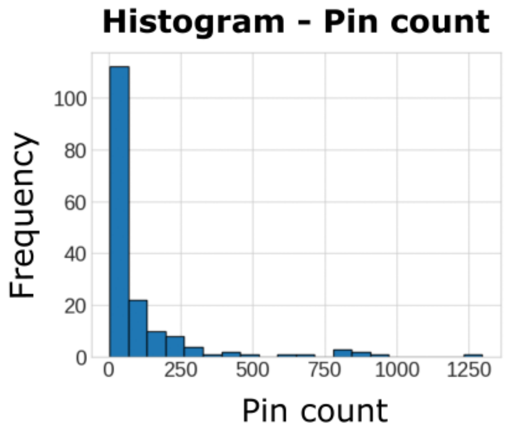








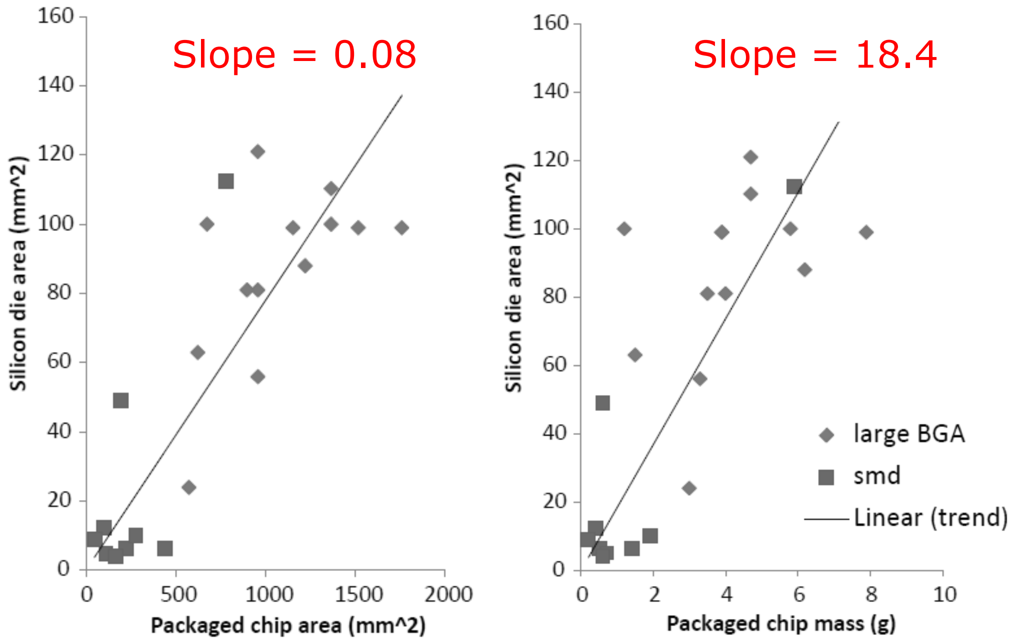




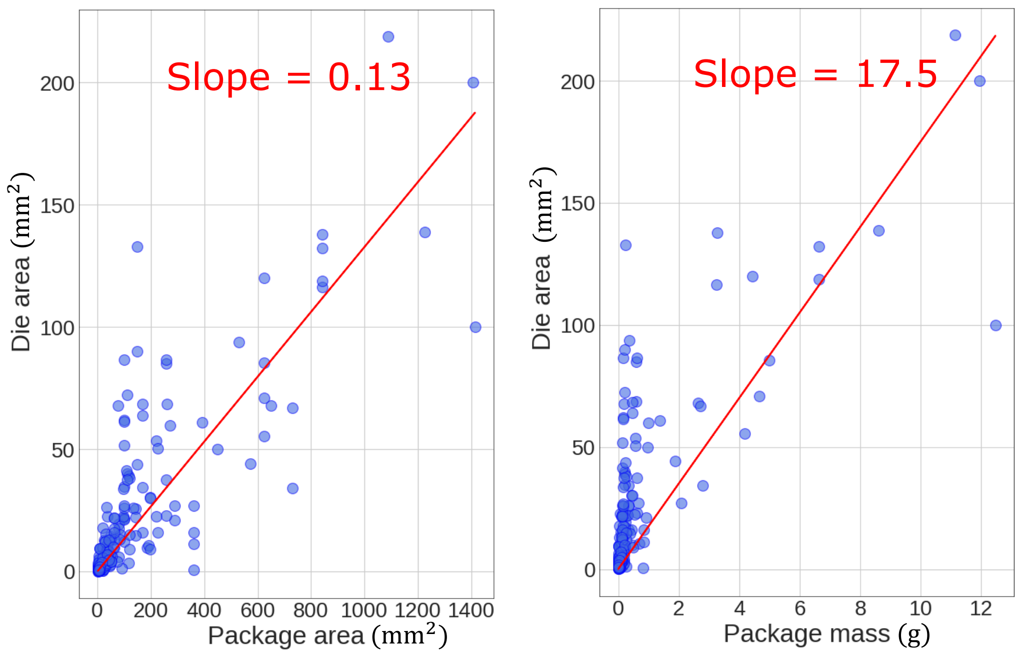
3. Parametric modeling

II. Linear univariate models

Figures from [1]



This work



[1] Teehan, P. (2014). Integrative approaches to environmental life cycle assessment of consumer electronics and connected media (Doctoral dissertation, University of British Columbia).

3. Parametric modeling

III. Linear and non-linear multivariate models

Performance metric used in this work: $R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$

3. Parametric modeling

III. Linear and non-linear multivariate models

Performance metric used in this work: $R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$

	Models	1.A – Linear [Teehan, 2014]	1.B – Linear [Teehan, 2014]	1.C - Linear	1.D - Linear	1.E - Linear
Features	Package area	×		×	×	×
	Package mass		×	×	×	×
	Pin count				×	×
	Package type + Function					×
	R^2 (train)	0.56	-0.63	0.58	0.58	0.64

3. Parametric modeling

III. Linear and non-linear multivariate models

Performance metric used in this work: $R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$

	Models	1.A – Linear [Teehan, 2014]	1.B – Linear [Teehan, 2014]	1.C - Linear	1.D - Linear	1.E - Linear
Features	Package area	×		×	×	×
	Package mass		×	×	×	×
	Pin count				×	×
	Package type + Function					×
	R^2 (train)	0.56	-0.63	0.58	0.58	0.64

	Models	2.A – MLP (100-100)	2.B – MLP (100-100)	2.D – KNN (K=2)	2.E – KNN (K=2)
Features	Package area	×	×	×	×
	Package mass	×	×	×	×
	Pin count	×	×		×
	Package type + Function		×		×
	R^2 (train)	0,41	0,59	0,86	0,85

3. Parametric modeling

IV. Models comparison

The objective of K-fold cross-validation is to reliably evaluate a model's performance by **training** and **testing** it **across multiple data splits**.

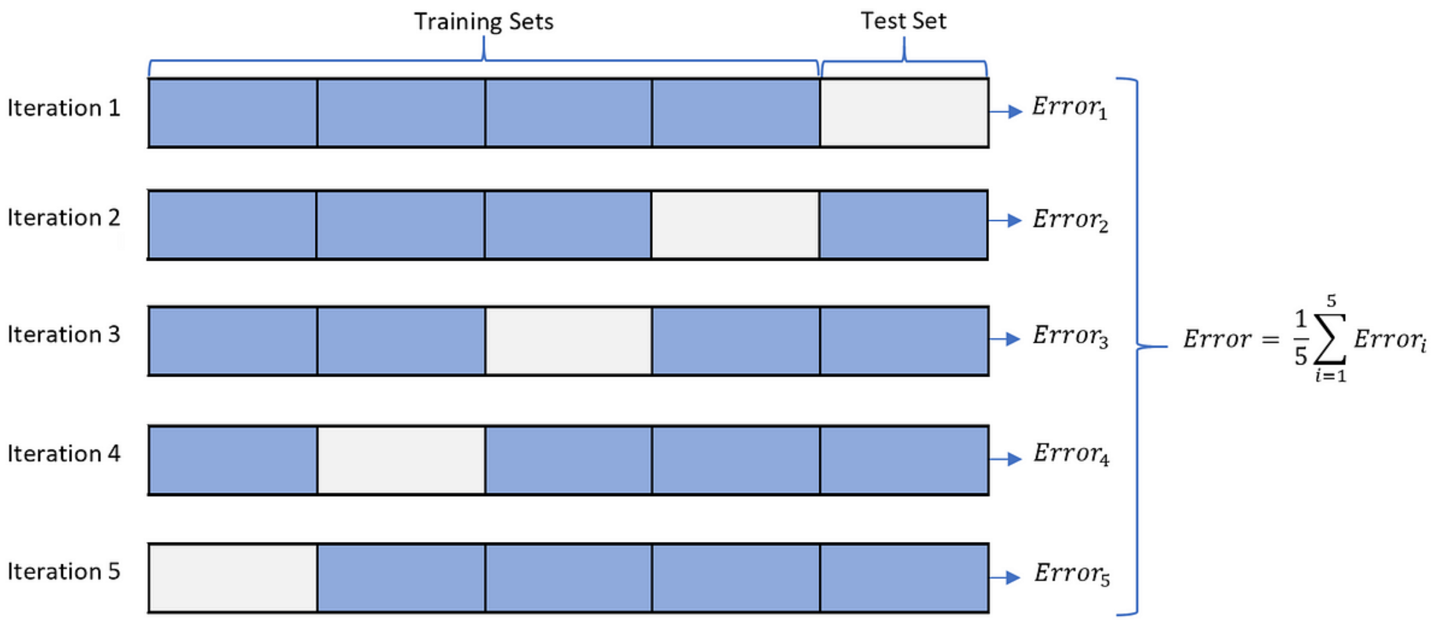
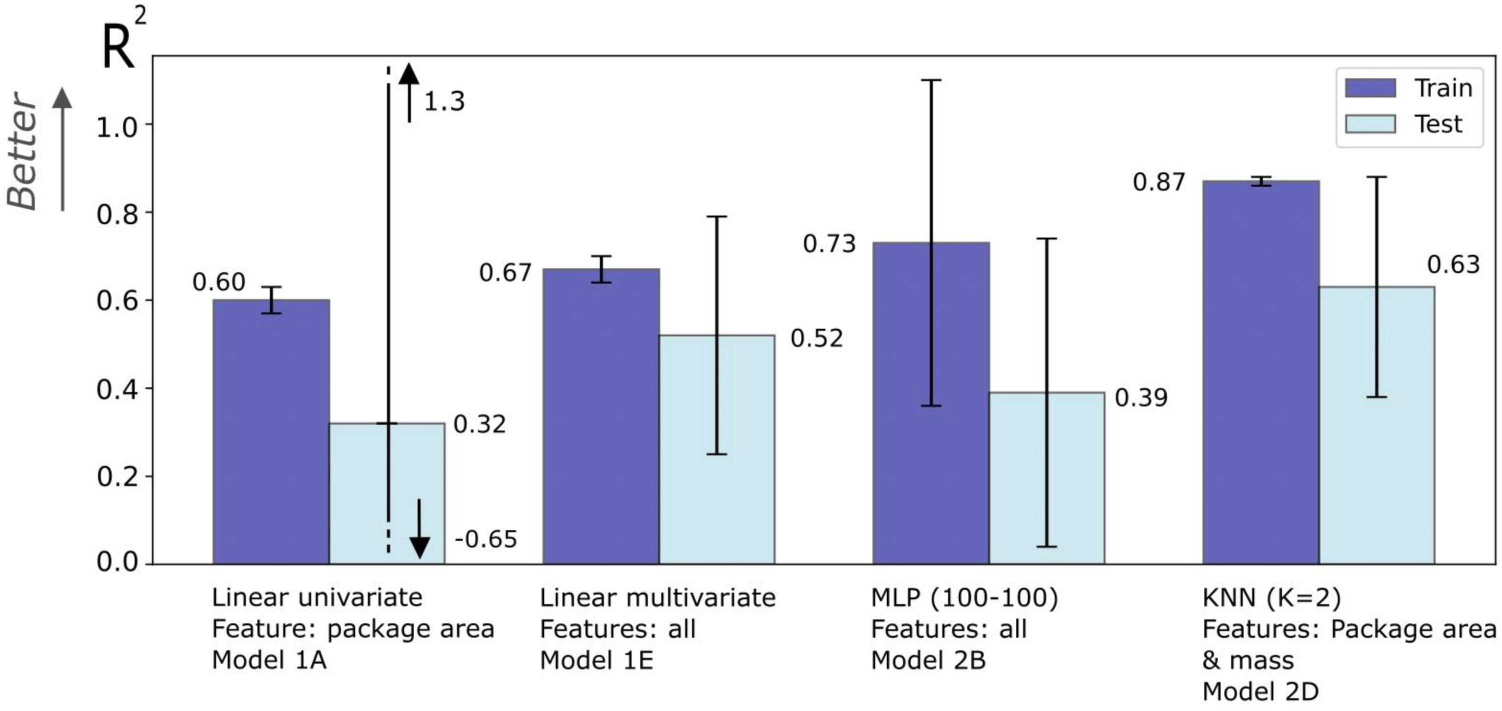


Illustration of K-fold cross validation

3. Parametric modeling

IV. Models comparison

Each data is used once in the test set (10% of the data).



4. Conclusions and future works

Can the die area be better estimated based on an extended feature set?

- **Univariate** linear model based solely on **package mass** exhibits the worst performance. $\mathbf{R^2}$ = - 0.62 (training).
  Common practices introduce significant uncertainty in LCA results.
- A **multivariate** linear model provides better predictions w.r.t. a **univariate** linear model, but it fails to combine all features efficiently.

Can other models be used to improve prediction?

- The use of a **KNN-type model (K=2)** provides the best performance so far. $\mathbf{R^2}$ = 0.63 $\pm$ 0.25.
  The model only relies on package area and mass features. KNN performs better if categorical features are not included.
- **MLP could probably achieve better performance**, as it would be possible to integrate **categorical features** more effectively. Due to the intrinsic characteristics of the model, the lack of data is a limiting factor.

4. Conclusions and future works

Future works

1. Gather **more data** to assess conclusion’s robustness.
2. **Extensive review** of **other models** that may prove more robust

   + **fine-tuning** of **current models** used to date.
3. **Open-source** publication of data and models

5. Taking a step back

I. Should we only rely on prediction models ?

- **The largest ICs of a computer** contribute to **30 to 70%** of the total silicon area [2].
- It would seem reasonable to:

  (1) Rely on models to estimate the die area of small-to-medium ICs.

  (2) Rely on measurements for the ICs larger than a given threshold.

→ Better modeling time vs. uncertainty trade-offs ?

II. Couldn't manufacturers provide the die area in IC datasheets?

- Currently, models are not performing well in predicting the die area of an IC.

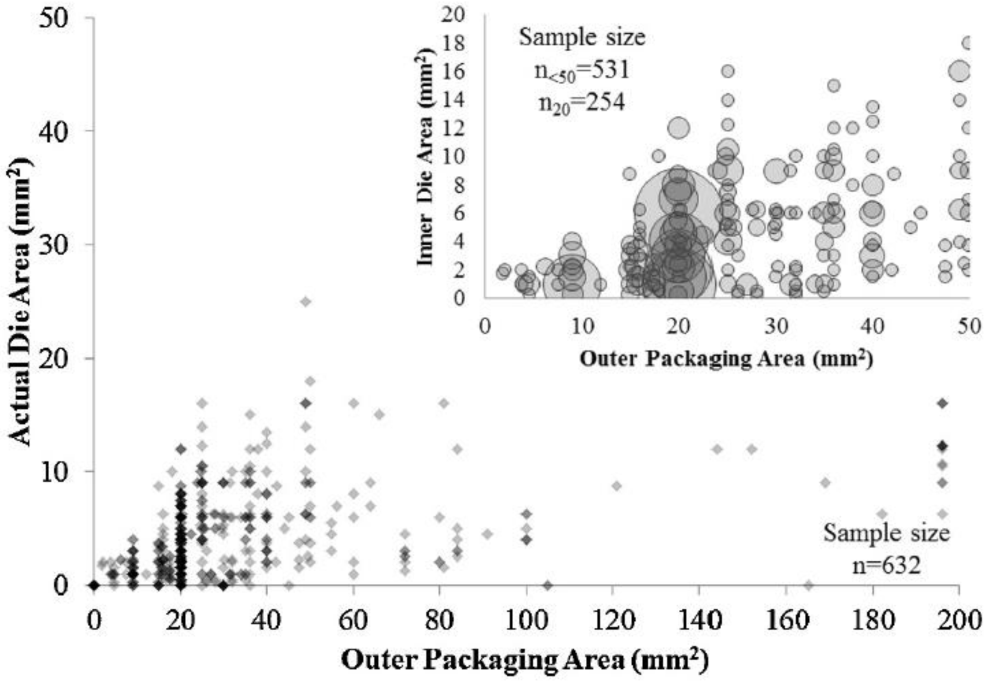
  Moreover, as stacked dies are becoming more common, these models will soon become outdated.

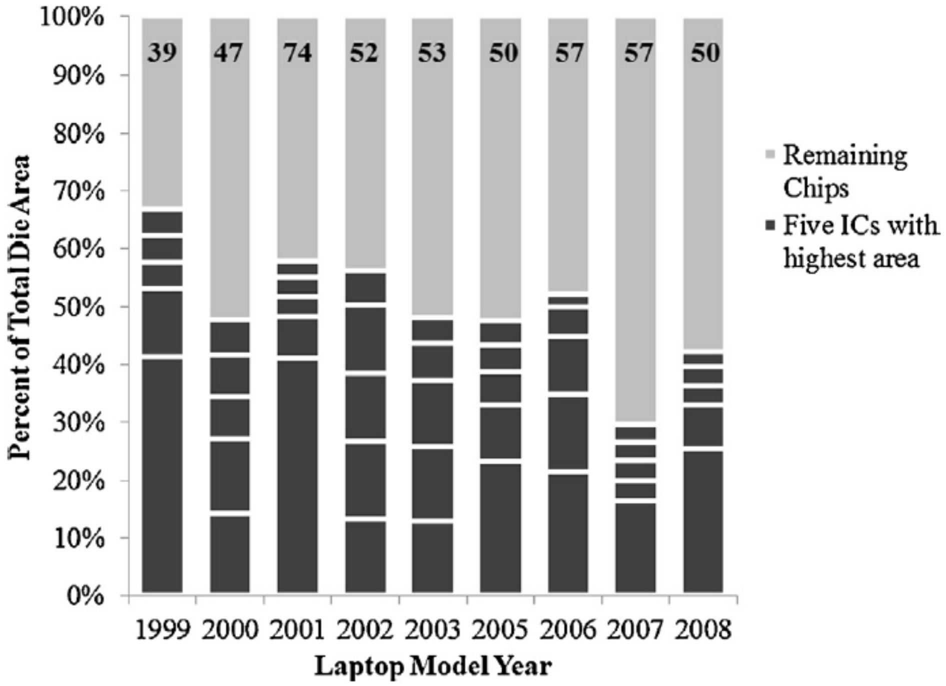
  It is important for manufacturers to **include the die area** of an IC in its **datasheet**.

[2] B. V. Kasulaitis, C. W. Babbitt, R. Kahhat, E. Williams, and E. G. Ryen, "Evolving materials, attributes, and functionality in consumer electronics: Case study of laptop computers," Resour. Conserv. Recycl., vol. 100, pp. 1–10, Jul. 2015.

Additional Slides

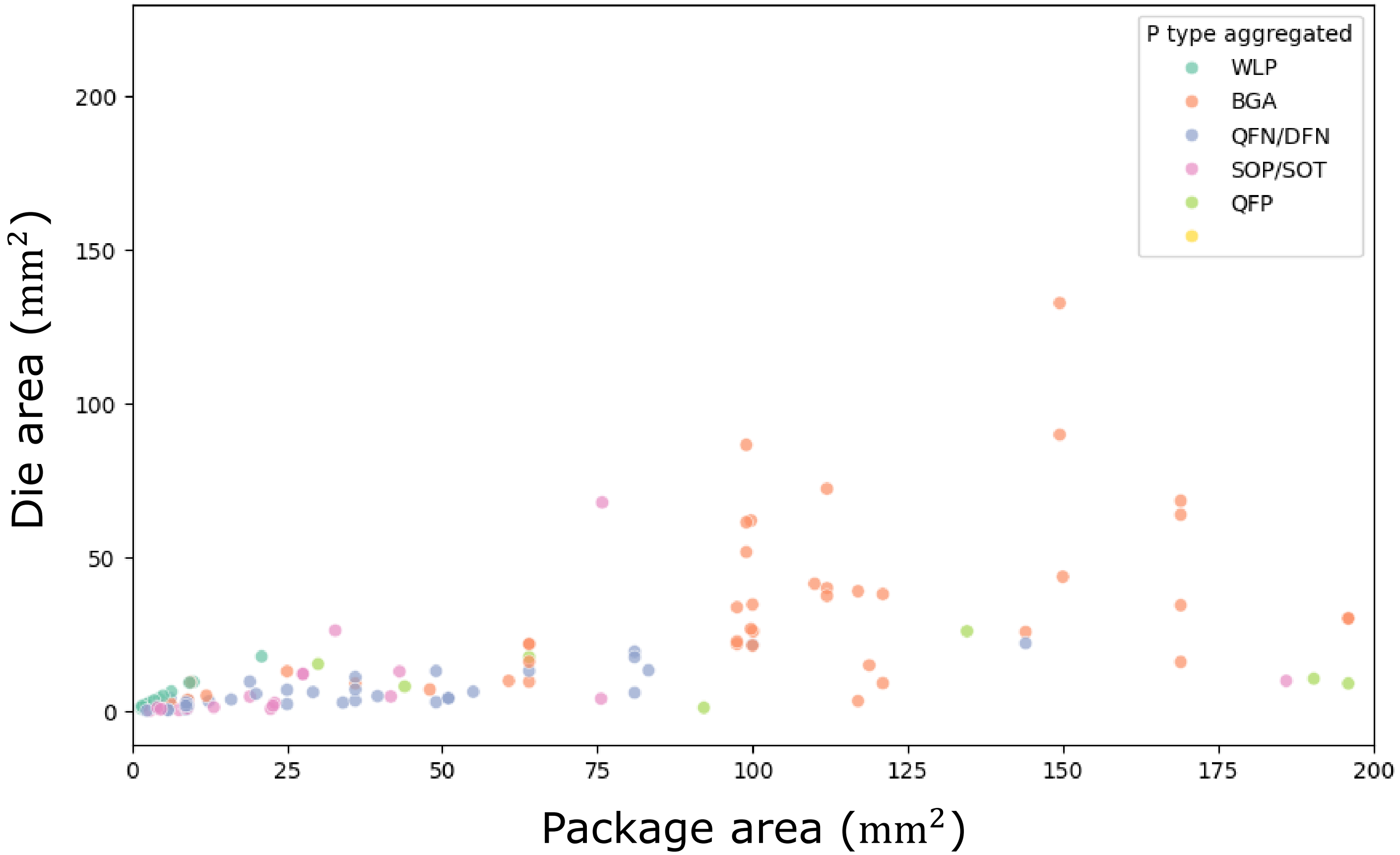
Kasulaitis's paper





Additional Slides

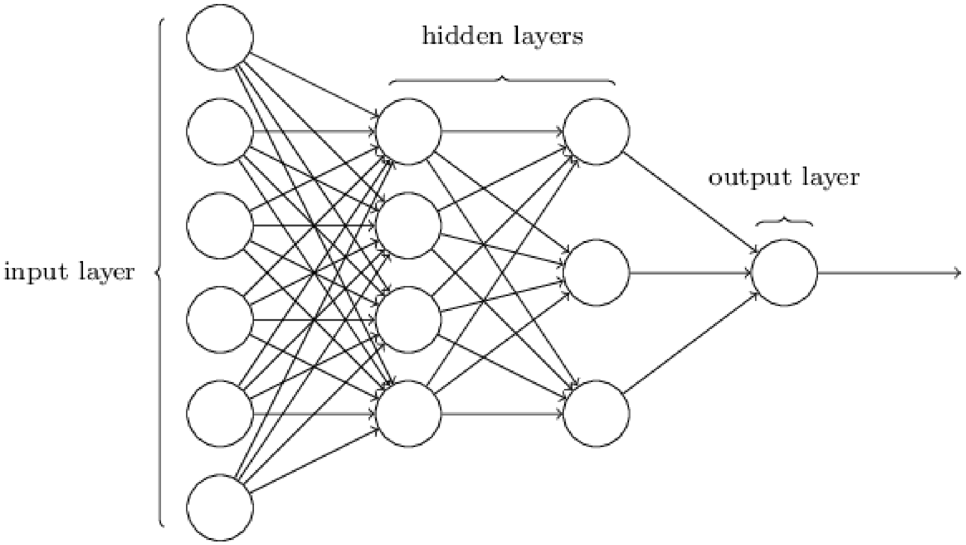
Heuristic-based models



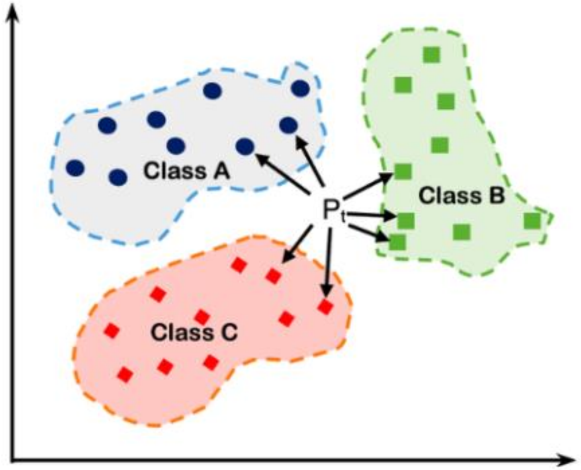
Additional Slides

Non-linear models – Description

Multilayer perceptron (MLP)

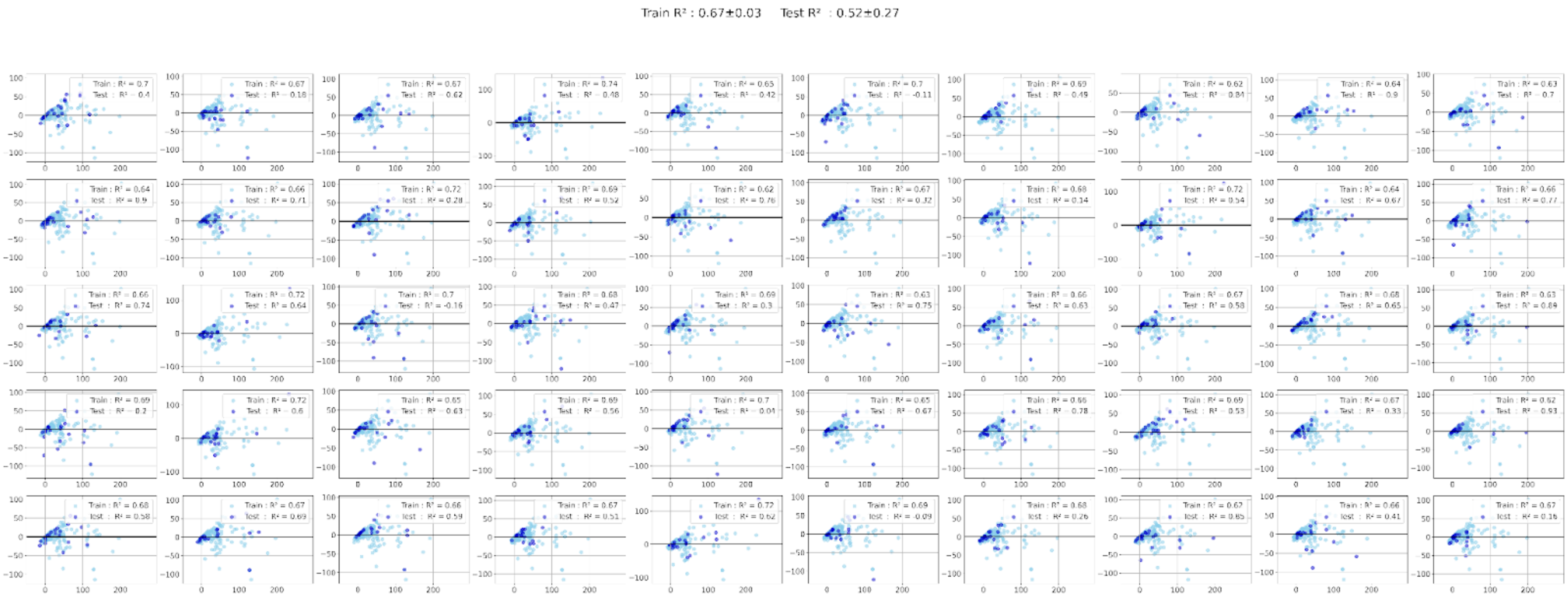


K nearest neighbors



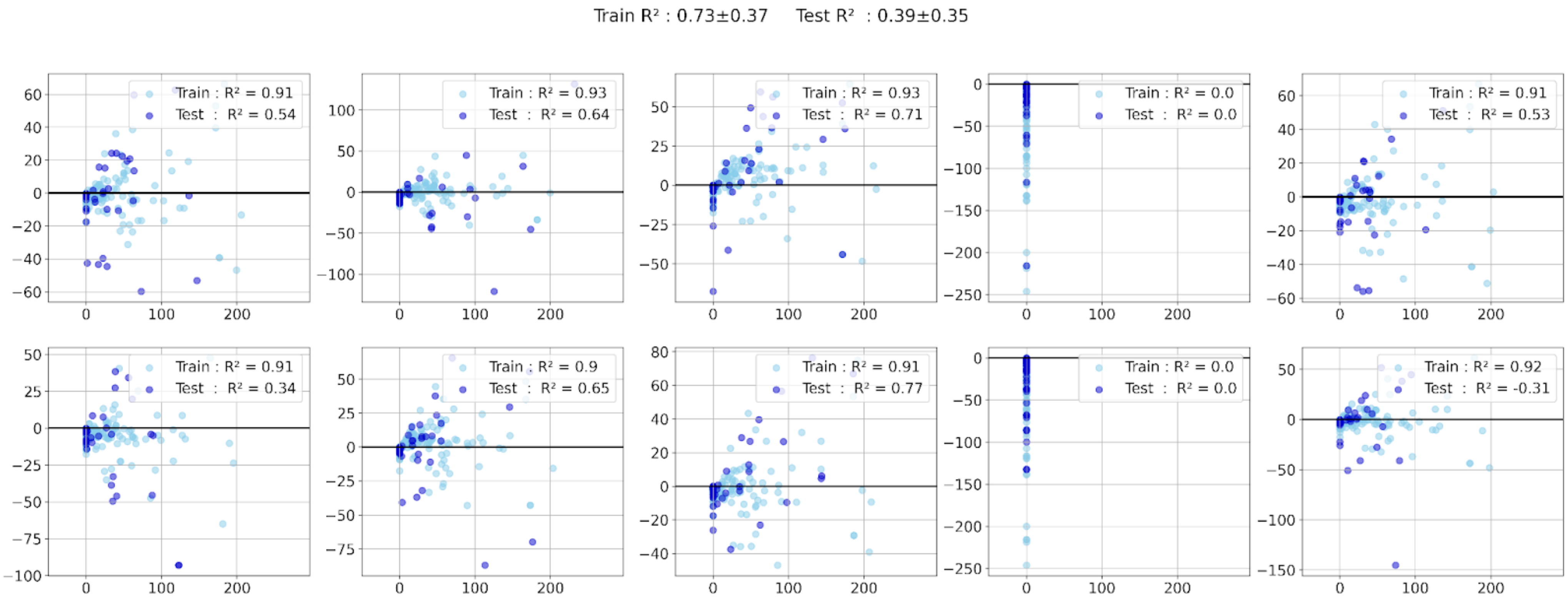
Additional Slides

Linear model – all features



Additional Slides

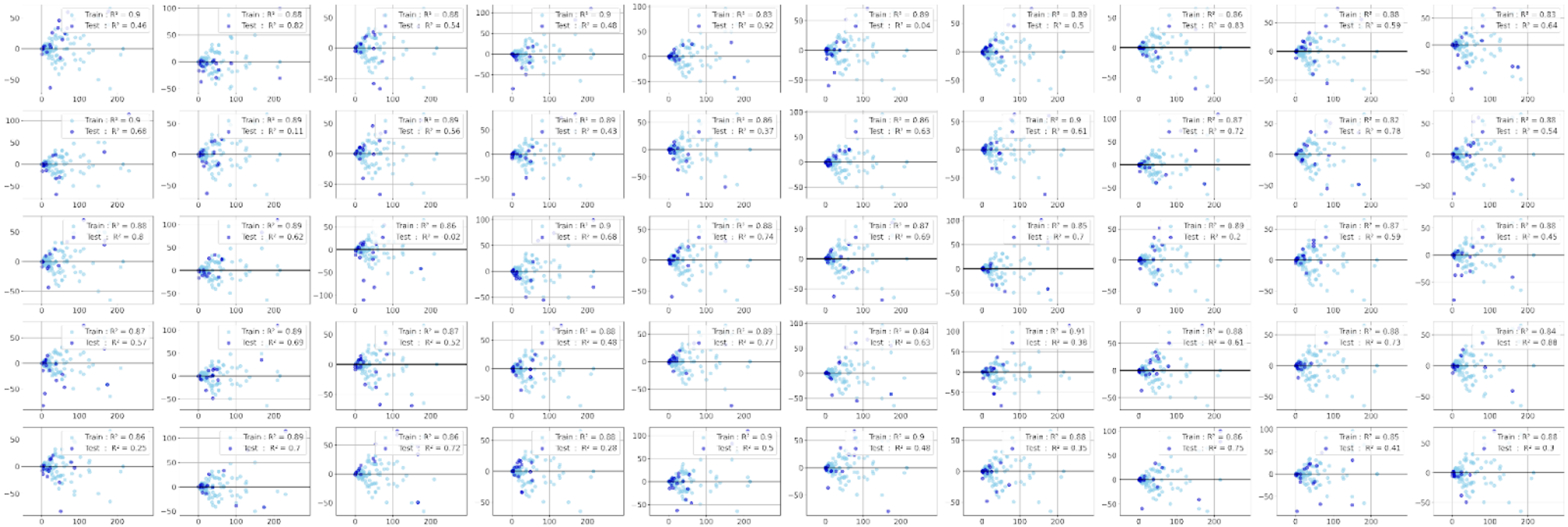
MLP – all features



Additional Slides

KNN – all features

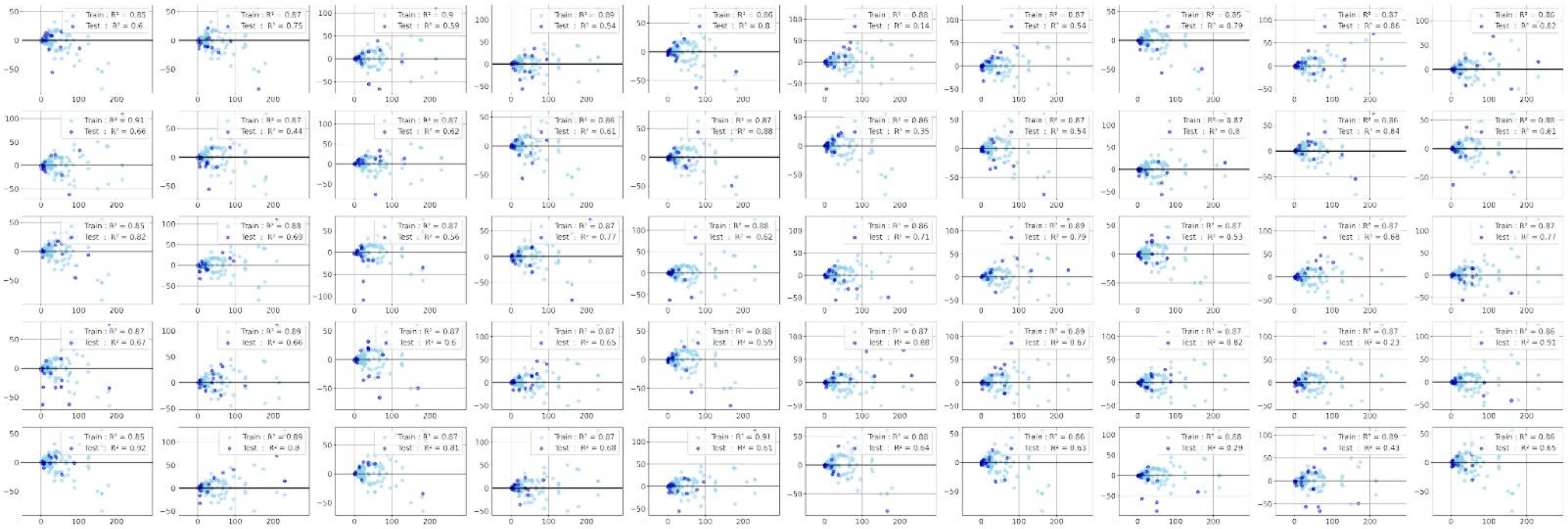
Train R^2 : 0.88±0.02   Test R^2 : 0.55±0.21



Additional Slides

KNN – Only Package area and mass

Train R^2 : 0.87±0.01   Test R^2 : 0.63±0.25



Additional Slides

SVM – poly kernel, 2d degree

Train R^2 : 0.64±0.04   Test R^2 : 0.33±0.42