

Documento: Descripción del Dataset WM-811K Wafer Map

Resumen General

El **WM-811K Wafer Map Dataset** es uno de los conjuntos de datos más extensos y detallados disponibles públicamente para el análisis de defectos en la fabricación de semiconductores. Contiene **811,457 mapas de obleas (wafers)**, recopilados de procesos de fabricación reales, y es ampliamente utilizado para entrenar y evaluar modelos de clasificación de patrones de defectos en obleas.

Estructura del Dataset

- **Total de muestras:** 811,457 mapas de obleas.
- **Etiquetas disponibles:** 172,950 muestras etiquetadas manualmente.
- **Etiquetas de patrones de defectos:**
 - **0 - Center:** 4,294 muestras (2.5%)
 - **1 - Donut:** 555 muestras (0.3%)
 - **2 - Edge-Loc:** 5,189 muestras (3.0%)
 - **3 - Edge-Ring:** 9,680 muestras (5.6%)
 - **4 - Loc:** 3,593 muestras (2.1%)
 - **5 - Random:** 866 muestras (0.5%)
 - **6 - Scratch:** 1,193 muestras (0.7%)
 - **7 - Near-full:** 149 muestras (0.1%)
 - **8 - None:** 147,431 muestras (85.2%)

Nota: La mayoría de las muestras etiquetadas pertenecen a la clase "None", lo que indica una distribución de clases altamente desbalanceada .

Detalles Técnicos

- **Formato de los mapas:** Imágenes en escala de grises de 256×256 píxeles.
 - **0:** Áreas fuera de la oblea.
 - **127:** Dies (chips) sin defectos.
 - **255:** Dies con defectos.
- **Columnas adicionales:**
 - **waferMap:** Mapa 2D de la oblea.
 - **dieSize:** Tamaño del die (chip) dentro de la oblea.
 - **lotName:** Nombre del lote de fabricación.

- waferIndex: Índice de la oblea dentro del lote.
- trainTestLabel: Indica si la muestra es para entrenamiento o prueba.
- failureType: Tipo de defecto (o "none" si no hay fallas).

Aplicaciones y Relevancia

- **Clasificación de patrones de defectos:** Utilizado para entrenar modelos que identifican y clasifican patrones de defectos en obleas.
- **Mejora del rendimiento de fabricación:** Ayuda a identificar causas raíz de defectos, mejorando el rendimiento y reduciendo costos.
- **Investigación académica:** Base para numerosos estudios en el campo de la detección de defectos y aprendizaje automático en la industria de semiconductores .

Recursos y Descargas

- **Repositorio en Kaggle:** <https://www.kaggle.com/datasets/qingyi/wm811k-wafer-map>
- **Repositorio en GitHub:** [WM-811K Semiconductor Wafer Map Pattern Classified](#)