

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02- 2020

Optimización de celdas solares basadas en perovskita: Desempeño comparativo de FAPbI_3 y MASnI_3 empleando técnicas de aprendizaje de máquinas

Jefferson Andrés Echeverri Pérez

Trabajo de grado presentado como requisito parcial para optar al título de:
Ingeniero en Electrónica

Asesor(es)
Ph.D. Erick Reyes Vera
Ph.D. Esteban González Valencia

Instituto Tecnológico Metropolitano - ITM
Facultad de Ingenierías
Departamento Electrónica y Telecomunicaciones
Medellín, Colombia
2025

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

RESUMEN

La transición global hacia fuentes de energía renovables exige soluciones fotovoltaicas que combinen alta eficiencia, bajo costo de fabricación y mínimo impacto ambiental. En este contexto, las celdas solares, y en especial las celdas solares de perovskita destacan por su rápido avance en la última década, debido a que estas han alcanzado eficiencias cercanas a las de las obtenidas con celdas de silicio. Además, gracias a su composición cristalina en la que tres elementos distintos se organizan de manera precisa, lo que permite modular con gran exactitud sus propiedades ópticas y eléctricas. Sin embargo, persisten desafíos críticos relacionados con la toxicidad del plomo en perovskitas tales como en la FAPbI_3 y la inestabilidad química de alternativas libres de plomo como MASnI_3 , cuya tendencia a oxidarse limita su desempeño.

Esta tesis presenta una metodología integral para optimizar celdas solares de perovskita tanto con FAPbI_3 (con plomo) como con MASnI_3 (sin plomo), combinando simulaciones numéricas avanzadas en SCAPS-1D y técnicas de aprendizaje de máquinas, específicamente aplicando el modelo Random Forest. Primero, se modelaron las celdas con arquitectura *n-i-p*, definiendo parámetros detallados de espesor, energía de banda prohibida, afinidad electrónica y densidades de defectos para cada capa. Empleando SCAPS-1D se generaron 1.000 configuraciones por cada tipo de perovskita, variando los espesores de ETL, capa activa y HTL para obtener las curvas J–V y calcular los indicadores eléctricos clave tales como eficiencia de conversión de potencia (PCE) y factor de llenado (FF).

A continuación, se desarrollaron dos modelos de Random Forest, el primero para predecir el FF y el segundo para estimar la PCE en función de los espesores. La validación cruzada y la búsqueda en malla de hiperparámetros garantizaron un ajuste óptimo. El modelo demostró alta precisión tanto de FF como de PCE, con errores relativos menores al 2 % en la mayoría de las configuraciones. Además, se creó una calculadora interactiva en Python que, mediante los modelos entrenados, permite estimar instantáneamente FF y PCE para cualquier combinación de espesores dentro del rango estudiado, evitando simulaciones repetitivas.

Mediante optimización multiobjetivo se construyó el frente de Pareto, identificando las configuraciones que maximizan simultáneamente PCE y FF sin sacrificar uno por el otro. Los resultados evidencian que las celdas con FAPbI_3 alcanzan eficiencias superiores (hasta $\approx 26\%$) y FF elevados ($\approx 83\%$), mientras que las basadas en MASnI_3 , aunque ligeramente menores en desempeño (PCE $\approx 25\%$, FF $\approx 82\%$), constituyen una alternativa más sostenible al eliminar el plomo. La comparación con datos experimentales y simulados publicados en la literatura confirma la capacidad del modelo para generalizar y predecir con precisión entornos reales. De esta manera, los hallazgos hallados en este trabajo de grado promueven el desarrollo de dispositivos fotovoltaicos más eficientes y amigables con el medio ambiente, sentando bases sólidas para futuras investigaciones en perovskitas libres de plomo y en estrategias de optimización híbrida.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02- 2020

Palabras clave: celdas solares de perovskita, FAPbI_3 , MASnI_3 , SCAPS-1D, aprendizaje automático, Random Forest, optimización multiobjetivo, PCE, FF, sostenibilidad.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

RECONOCIMIENTOS

Agradezco principalmente a mi familia, que ante las adversidades que la vida nos ha puesto, siempre ha existido el apoyo sobre todas las decisiones que han sido tomadas en el transcurso de toda mi carrera. Además, agradezco a mis asesores, Erick Reyes Vera y Esteban González Valencia, profesionales que no dudaron de mi proceso y capacidades académicas, profesionales y personales para poder culminar con excelencia este proyecto.

Quiero agradecer a todos mis compañeros que me han acompañado en este proceso académico, a Yeraldin Vélez, quien me ha enseñado con paciencia todo sobre el potencial que este trabajo tiene, Sebastián Valencia, quien amablemente ha sido un gran asesor en el manejo de diferentes softwares; y al resto de compañeros que han aportado algo de su tiempo para la solución de dudas que han surgido a lo largo de la dedicación de este proyecto.

Este proyecto lo dedico a mi madre y a mi padre, Sandra Milena Pérez y Geovanny Andrés Echeverri que han conocido todo mi proceso académico y personal y nunca han dudado de mis capacidades, que saben lo valioso que es para mí siempre contar con el apoyo de ambos en mi formación profesional.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

ACRÓNIMOS


FTO: Oxido de estaño dopado con flúor
SnO₂: Dióxido de estaño
TiO₂: Dióxido de titanio
ETL: Capa de transporte de electrones
PVK: Capa fotovoltaica de Perovskita
HTL: Capa de transporte de huecos
Au: símbolo de oro (contacto metálico de celda solar)
FAPbI₃: formamidinio plomo yoduro
MASnI₃: metilamonio estaño yoduro
FF: Facto de llenado
Voc: Voltaje de circuito abierto
Jsc: Corriente de corto circuito
PCE: eficiencia de conversión de potencia
SCAPS-1D: Solar Cell Capacitance Simulator
IR: Infrarrojo
ML: Machine Learning o Aprendizaje de máquina
RF: Random Forest o Bosque Aleatorio
XGB o XGBoost: Extreme Gradient Boosting
SVR: Support Vector Regression
NN: Neural Networks o Redes Neuronales
LR: Linear Regression o Regresión Lineal
DFT: Density Functional Theory o Teoría del Funcional de la Densidad
CART: Classification and Regression Trees
RSS: Residual Sum of Squares
KI: Yoduro de potasio (Potassium Iodide)
MAPbI₃: Metilamonio plomo yoduro
FASnI₃: Formamidinio yoduro de estaño
TCAD: Technology Computer-Aided Design
COMSOL: COMSOL Multiphysics
Spyder: IDE para Python, aunque no es un acrónimo técnico, podrías aclararlo como entorno de desarrollo
AHK: AutoHotKey, lenguaje de automatización para Windows
AM1.5G: Estándar del espectro solar Air Mass 1.5 Global
J–V: Current density–Voltage curve o Curva de densidad de corriente–voltaje

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

TABLA DE CONTENIDO

Contenido

1 INTRODUCCION.....	9
Objetivos generales.....	10
Objetivos específicos	10
2 MARCO TEORICO.....	12
2.1 Principios fundamentales de las celdas solares de perovskita.....	11
Arquitecturas <i>n-i-p</i> y <i>p-i-n</i>	11
Materiales de las capas en la arquitectura <i>n-i-p</i>	13
Capa de transporte de electrones (ETL)	13
Capa activa de perovskita.....	14
Capa de transporte de huecos (HTL).....	15
2.2 Parámetros eléctricos y ópticos en celdas solares	16
2.3 Herramientas de simulación para dispositivos fotovoltaicos.	22
2.4 Aprendizaje de máquina en la optimización de celdas solares.....	24
Revisión general de tareas de regresión en sistemas físicos complejos.....	24
Explicación y estructura del modelo: arboles de Clasificación y Regresión (CART)..	25
Desventajas de los árboles de decisión.....	26
Bosques aleatorios: fundamentos y construcción.....	27
Aleatorización: doble fuente de variabilidad.....	27
Relación entre varianza y correlación	28
Construcción del bosque.....	28
Medida de importancia por permutación.....	28
Selección de variables con Random Forest	28
Impacto de la correlación.	29
Entrenamiento del modelo: RandomForestRegressor	29
Métricas de desempeño	30
Coeficiente de determinación (R^2).....	30
Error Absoluto Medio (MAE).....	30

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Raíz del error cuadrático medio (RMSE).....	31
Optimización mediante hiperparámetros	31
2.5 Técnicas de optimización multiobjetivo.....	32
Concepto de Frente de Pareto.....	32
Maximización ponderada.....	33
3 METODOLOGIA.....	36
3.1 Modelamiento de celdas solares de perovskita mediante SCAPS-1D	35
3.1.1 Diseño de celdas solar de perovskita de múltiples capas.....	35
3.1.2 Simulación de propiedades ópticas y recombinación	36
3.2 Metodología de simulación con el uso de SCAPS-1D.....	38
3.3 Extracción y tratamiento de datos	40
Generación de simulaciones en modo Batch.....	41
Manejo de errores y validación con AutoHotKey (AHK).....	41
Extracción de resultados mediante Python.....	42
3.4 Modelo de optimización mediante aprendizaje de máquina	42
3.4.1 Algoritmo Random Forest para optimización de celda solar.....	44
Optimización multiobjetivo mediante frente de Pareto.....	45
Maximización ponderada de FF y PCE.....	46
Herramienta interactiva de estimación	47
Optimización de hiperparámetros del modelo de ML.....	48
4 RESULTADOS Y DISCUSIÓN.....	52
4.1 Resultados mediante técnicas multiobjetivo.....	49
4.2 Análisis de los resultados para celdas solares de FAPbI ₃	50
4.3. Análisis de los resultados para celdas solares de MASnI ₃	52
4.4. Comparación del rendimiento entre celdas solares basadas en FAPbI ₃ y MASnI ₃	54
4.5. Capacidad de generalización del modelo y comparación con resultados de la literatura	54
4.6. Comparación del rendimiento de celdas solares MASnI ₃ y evaluación del modelo predictivo.....	56
4.7. Desempeño del modelo de predicción.....	57
Metricas de evaluacion para celda solar FAPbI ₃	57
Metricas de evaluacion para celda solar MASnI ₃	57
5 CONCLUSIONES Y RECOMENDACIONES.....	63
REFERENCIAS	61

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

1. INTRODUCCIÓN

La transición hacia fuentes de energía renovables es una de las prioridades globales para reducir la dependencia de los combustibles fósiles y mitigar los efectos del cambio climático. Para ello, se han explorado diversas alternativas con miras a generar energías limpias, tales como la generación de energía eólica, mareomotriz, hidrogeno verde, solar entre otras. Dentro de las tecnologías disponibles, la energía solar ha demostrado ser una de las opciones más prometedoras, gracias a la abundancia del recurso solar y a los avances en la eficiencia de conversión de las celdas fotovoltaicas. Hoy en día, las celdas solares de silicio dominan el mercado, alcanzado eficiencias récord de hasta 27.6% en pruebas de laboratorio[1]. Aunque en sus inicios la producción de estas celdas implicaba costos elevados y procesos energéticamente intensivos, el desarrollo industrial y la economía de escala han permitido una notable disminución de los costos. Según datos recientes, el precio de los paneles solares de silicio ha caído aproximadamente un 90% en los últimos años, haciendo que la energía solar sea hoy más accesible que nunca. Sin embargo, desafíos como la rigidez estructural del silicio aun limitan su aplicación en ciertos contextos, lo que ha impulsado la investigación en tecnologías emergentes como las celdas solares de perovskita, más ligeras, flexibles y potencialmente más baratas [2].

En este escenario, las celdas solares de perovskita han emergido como una tecnología disruptiva en el campo de la energía fotovoltaica, debido a su alta eficiencia de conversión, bajo costo de fabricación y facilidad de procesamiento mediante técnicas de deposición en solución y escalada industrial [3], [4]. Desde su introducción en 2009, la eficiencia de las celdas de perovskita ha experimentado un crecimiento exponencial, alcanzando valores superiores a la eficiencia de las celdas solares de silicio en menos de una década [5], demostrando su potencial en esta área. Además, cuando se combinan con silicio en una configuración tándem, estas celdas han logrado superar el 33.24% de eficiencia, lo que las posiciona como una opción clave para el futuro de la generación fotovoltaica [6].

Las perovskitas se describen con la fórmula general ABX_3 , donde A es un catión monovalente (orgánico o inorgánico), B es un catión metálico bivalente (por ejemplo, Pb^{2+} o Sn^{2+}) y X es un anión haluro (I^- , Br^- o Cl^-). Esta red tridimensional ABX_3 permite intercambiar casi sin límites los iones A, B y X, ajustando sus propiedades ópticas, eléctricas y estructurales de forma muy precisa [7][8].

Originalmente, las perovskitas se emplearon como material fotosensibilizador en celdas solares sensibilizadas con colorante (DSSC), en las que la película de perovskita sustituyó al tinte tradicional para captar la luz. Con el tiempo, estas estructuras han evolucionado rápidamente hasta convertirse en el núcleo de la tercera generación de dispositivos fotovoltaicos [4]. Esta rápida evolución ha sido posible gracias a la optimización de numerosos parámetros tales como la calidad del cristal, la ingeniería de interfaces y la

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

composición química. Entre las perovskitas, algunas de las más comunes incluyen plomo en su estructura, como lo son el MAPbI_3 (metilamonio yoduro de plomo), FAPbI_3 (formamidinio yoduro de plomo) y CsPbI_3 (cesio yoduro de plomo). Estos compuestos han mostrado alta eficiencia. Sin embargo, la toxicidad del plomo y su impacto ambiental han generado preocupación, limitando la viabilidad comercial de estas celdas y promoviendo la exploración de alternativas más seguras. Para abordar esta problemática, se han investigado el uso de perovskitas libres de plomo, siendo las más destacadas MASnI_3 (metilamonio yoduro de estaño) y FASnI_3 (formamidinio yoduro de estaño), en las cuales se sustituye el plomo por estaño. Las perovskitas de estaño tienen propiedades optoelectrónicas similares a las de plomo, pero enfrentan problemas adicionales relacionados con su estabilidad química y su tendencia a oxidarse.

Uno de los enfoques más prometedores en la optimización de celdas solares de perovskita es la aplicación de técnicas de aprendizaje de máquinas (ML, por sus siglas en inglés) [9]. Estas herramientas han revolucionado la investigación en materiales fotovoltaicos, al permitir examinar grandes volúmenes de datos experimentales y simulaciones computacionales con el fin de identificar patrones complejos en la relación entre estructura y desempeño. Gracias a estos avances, el ML se ha convertido en una estrategia clave para predecir la eficiencia de conversión, optimizar la composición de las perovskitas y reducir la cantidad de ensayos físicos necesarios en el desarrollo de nuevos materiales. Varios estudios han demostrado que los modelos de ML pueden emplearse para clasificar perovskitas en función de su estabilidad, eficiencia y propiedades ópticas, lo que acelera la búsqueda de materiales con mayor rendimiento fotovoltaico. Más allá de la simple clasificación de perovskitas, hoy en día se emplean modelos de bosque aleatorio (Random Forest), gradiente potenciado (XGBoost), redes neuronales artificiales (RNA) y regresión lineal para predecir con alta precisión parámetros críticos como la eficiencia de conversión de potencia (PCE) y la banda prohibida de nuevos compuestos [10]. En estudios de diseño inverso, se han combinado conjuntos de datos de alta fidelidad procedentes de simulaciones DFT y datos experimentales con bosques aleatorios multilocalidad para ajustar la relación entre la composición ABX_3 y el rendimiento fotovoltaico, e integrarlos con algoritmos genéticos que exploran automáticamente miles de composiciones propuestas, obteniendo candidatos con PCE superiores al 15% y baja energía de descomposición [11].

Además, métodos de optimización bayesiana se utilizan para la calibración óptima de hiperparámetros de estos modelos y para seleccionar espesores y secuencias de capas transportadoras de carga, minimizando defectos interfaciales y maximizando la estabilidad del dispositivo [12]. De hecho, la optimización de las celdas solares mediante ML no se ha limitado a la predicción de nuevos materiales, sino que también ha permitido mejorar aspectos clave del diseño del dispositivo. Algoritmos de optimización, como los algoritmos genéticos (GA) y los modelos de optimización bayesiana también han sido aplicados para mejorar la selección de capas de transporte de carga, ajustar el espesor de las películas delgadas y reducir defectos en la interfaz entre capas [13].

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Esta tesis tiene como objetivo cerrar una de las principales brechas en el conocimiento al comparar el rendimiento de FAPbI_3 y MASnI_3 , con el fin de mejorar el diseño de las celdas solares de perovskita al ajustar los espesores de sus capas funcionales. Se creó un modelo que usa Random Forest, una técnica popular de aprendizaje de máquinas que ayuda a optimizar sistemas complejos, para encontrar las mejores condiciones de diseño estas celdas solares. El modelo se centra en la optimización del espesor de la capa de transporte de huecos (HTL), la capa de transporte de electrones (ETL) y la capa activa de perovskita. El objetivo es maximizar el factor de llenado (FF por su sigla en inglés) y la eficiencia de conversión de potencia. Para ello, se utilizaron datos de simulaciones computacionales generados en SCAPS-1D, los cuales permitieron entrenar el modelo y predecir combinaciones óptimas de espesores que conduzcan a un mejor rendimiento del dispositivo en función de los parámetros FF y PCE.

La estructura de esta tesis se organiza en cinco capítulos que permiten desarrollar el estudio de forma progresiva y coherente. El capítulo 1 introduce el contexto energético global y justifica el uso de perovskitas como alternativa a las tecnologías fotovoltaicas tradicionales, señalando la problemática ambiental del plomo y el potencial del estaño como sustituto. En el capítulo 2 se construye el marco teórico que sustenta la investigación, abordando desde los principios físicos de las celdas solares hasta las herramientas computacionales empleadas y los fundamentos del aprendizaje de máquinas aplicados a sistemas complejos. El capítulo 3 detalla la metodología seguida, que combina simulaciones numéricas mediante SCAPS-1D y técnicas de modelado predictivo con Random Forest para optimizar el diseño de las celdas. A continuación, en el capítulo 4 se presentan y analizan los resultados obtenidos, comparando el desempeño de los materiales FAPbI_3 y MASnI_3 en distintos escenarios de espesor de capa. Finalmente, el capítulo 5 resume las principales conclusiones, plantea recomendaciones técnicas y propone líneas futuras de investigación orientadas a la mejora de dispositivos fotovoltaicos sostenibles.

Objetivos generales

Optimizar celdas solares de Perovskita con y sin plomo, a través de la implementación de metodologías computacionales basadas en SCAPS-1D y la aplicación de técnicas de aprendizaje automático para maximizar el rendimiento eléctrico de estas

Objetivos específicos

- Implementar y validar una metodología computacional utilizando SCAPS-1D para modelar celdas solares de perovskita con y sin plomo, permitiendo la simulación precisa de su comportamiento bajo diferentes condiciones.
- Evaluar el desempeño de distintas celdas solares de perovskita, tanto con plomo como sin plomo, enfocándose en su rendimiento eléctrico y estabilidad a través de sus ajustes en los parámetros de diseño.
- Implementar un modelo de aprendizaje de máquinas para maximizar la eficiencia energética y el factor de llenado de las celdas solares de perovskita en función de los espesores de los materiales empleados en su diseño

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

2. MARCO TEÓRICO

En este capítulo se establecen los fundamentos conceptuales y metodológicos necesarios para comprender y abordar la optimización de celdas solares de perovskita mediante simulación computacional y técnicas de aprendizaje automático. En la sección 2.1 se describen los principios físicos de las celdas fotovoltaicas, incluyendo las arquitecturas $n-i-p$ y $p-i-n$, los mecanismos de generación de pares electrón-hueco y el transporte de carga en materiales semiconductores. La sección 2.2 profundiza en los parámetros eléctricos clave (V_{oc} , J_{sc} , FF y PCE) y en los conceptos ópticos esenciales coeficiente de absorción, índice de refracción y banda prohibida que determinan el desempeño de la celda. Seguidamente, en la sección 2.3, se realiza un análisis comparativo de las principales herramientas de simulación de dispositivos fotovoltaicos, destacando la eficiencia y versatilidad de SCAPS-1D para estudios paramétricos de capas delgadas., la sección 2.4 aborda la manera en que funciona el aprendizaje de máquina en la optimización de celdas solares introduciendo el algoritmo Random Forest para la optimización multiobjetivo de espesores de capa. Finalmente, la sección 2.5 introduce las técnicas de optimización multiobjetivo.

2.1 Principios fundamentales de las celdas solares de perovskita

En esta sección se establecen los conceptos físicos esenciales que permiten comprender el funcionamiento de las celdas solares basadas en perovskita. A lo largo de los apartados siguientes analizaremos, de manera progresiva y sistemática, la arquitectura multicapa típica, los procesos de generación y separación de pares electrón-hueco, los mecanismos de transporte y recolección de portadores, así como las principales vías de recombinación que limitan la eficiencia. De este modo, sentaremos las bases teóricas necesarias para interpretar los resultados de simulación y optimización descritos en los capítulos posteriores.

Arquitecturas $n-i-p$ y $p-i-n$

La estructura $n-i-p$ en celdas solares de perovskita consta de tres capas semiconductoras organizadas en la siguiente secuencia (véase Figura 2-1a):

- Capa tipo n (ETL): Esta capa esta dopada negativamente. La capa tipo n suele funcionar como una capa selectora de electrones, es decir, facilita el transporte de electrones mientras bloquea los huecos.
- Capa intrínseca (i): Esta capa del centro, es decir, corresponde a la capa de perovskita absorbente. En esta capa es donde ocurre la absorción fotónica de la luz solar y se generan los pares electrón-hueco
- Capa tipo p (HTL): Esta capa esta dopada positivamente para conducir huecos hacia el electrodo posterior y evitar la recombinación de electrones.

Esta disposición aprovecha el gradiente de concentración de portadores y el campo eléctrico interno para dirigir los electrones hacia el electrodo negativo y los huecos hacia el positivo, estableciendo así una corriente continua cuando el dispositivo se conecta a una carga externa.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Su uso en tecnologías como las celdas solares de perovskita ha sido clave por su compatibilidad con materiales de baja temperatura y su capacidad para ser procesada en capas delgadas, lo que la hace ideal para aplicaciones de bajo costo, flexibles o incluso semitransparentes [14].

Por otro lado, en la arquitectura $p-i-n$ las capas funcionales se disponen en orden inverso a la estructura $n-i-p$ (véase Figura 2-1b):

- Capa tipo p (HTL).
- Capa intrínseca absorbente i (perovskita).
- Capa tipo n (ETL).

En esta configuración la luz incide primero sobre la HTL (por ejemplo, PEDOT: PSS), atraviesa la capa de perovskita y llega finalmente a la ETL (por ejemplo, ZnO o SnO₂). En esta arquitectura se suelen emplear materiales compatibles con procesos a baja temperatura, como PEDOT: PSS (poli(3,4-etilendioxitiofeno): poliestirensulfonato) para el transporte de huecos, y óxidos semiconductores como ZnO o SnO₂ para el transporte de electrones. Esta configuración es habitual en celdas solares flexibles o impresas, donde se priorizan métodos de fabricación económicos y sustratos ligeros [15].

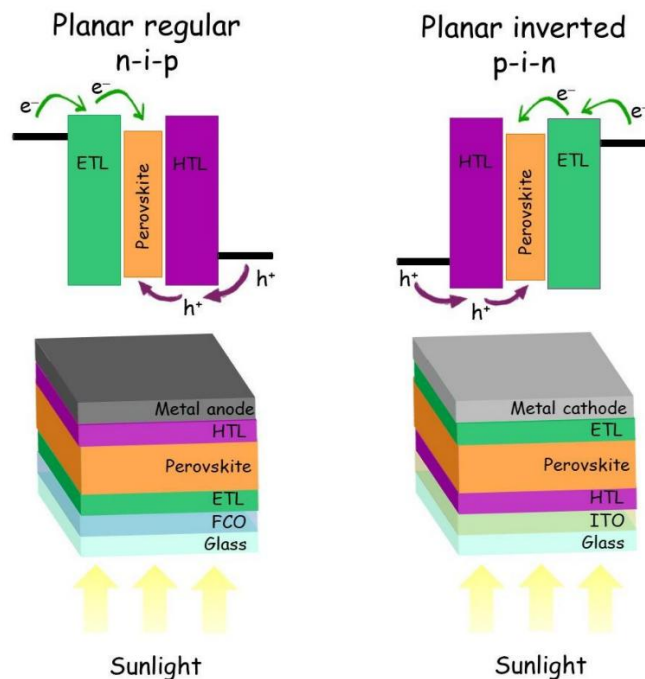


Figura 2-1. Esquema comparativo de las arquitecturas $n-i-p$ (a) y $p-i-n$ (b) en celdas solares de perovskita [14].

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Tabla 2-1. Comparación entre las arquitecturas tipo *n-i-p* y tipo *p-i-n*.

Característica	<i>n-i-p</i>	<i>p-i-n</i>
Orden de capas	$n \rightarrow i \rightarrow p$	$p \rightarrow i \rightarrow n$
Entrada de luz	Por el lado de la ETL	Por el lado de la HTL
Capa de transporte de huecos	HTL al final del dispositivo	HTL al inicio del dispositivo
Capa de transporte de electrones	ETL en la parte superior	ETL en la parte inferior
Electrodo frontal	FTO (óxido conductor transparente)	ITO (óxido conductor transparente)
Aplicaciones comunes	Celdas rígidas, vidrio	Celdas flexibles, impresión en solución
Procesamiento	Generalmente por deposición de alta temperatura	Compatible con técnicas de baja temperatura

La comparación entre las estructuras *n-i-p* y *p-i-n* permite identificar diferencias clave en su configuración, materiales y aplicaciones. La disposición de capas en cada arquitectura determina el orden de entrada de la luz y el trayecto de los portadores de carga, lo cual impacta en el rendimiento óptico del dispositivo [16].

En esta tesis se seleccionó la arquitectura *n-i-p*. En primer lugar, es la configuración más ampliamente reportada en la bibliografía para perovskitas como $FAPbI_3$ y $MASnI_3$, lo que facilita la comparación directa de resultados con trabajos previos y la validación de parámetros en SCAPS-1D. En segundo lugar, la combinación de óxidos inorgánicos TiO_2 como ETL y Spiro-OMeTAD como HTL aporta una mayor estabilidad térmica y eléctrica, permitiendo realizar barridos paramétricos extensos sin degradaciones prematuras. Por último, la simplicidad del flujo de procesos a través de spin-coating y tratamiento térmico controlado asegura una reproducibilidad elevada en las simulaciones que se realizaran, aspectos críticos para el entrenamiento y la validación del modelo de optimización propuesto.

Materiales de las capas en la arquitectura *n-i-p*

En esta sección detallamos los materiales más habituales para cada una de las capas que conforman la arquitectura *n-i-p* de las celdas solares de perovskita.

Capa de transporte de electrones (ETL)

TiO₂ (dióxido de titanio): El TiO_2 es quizás la capa de transporte de electrones más clásica, gracias a su excelente estabilidad térmica y una banda de conducción (~ 3.2 eV) bien alineada con la perovskita. Sin embargo, alcanzar la fase cristalina óptima exige temperaturas de sinterización superiores a 450 °C, lo que limita su compatibilidad con sustratos flexibles o sensibles al calor. Además, su movilidad electrónica del orden de ($1-10$ $cm^2/V \cdot s$) aunque adecuada, es inferior a la de óxidos como SnO_2 .

SnO₂ (óxido de estaño): El SnO_2 ofrece una elevada movilidad de electrones ($20-100$ $cm^2/V \cdot s$), además puede procesarse a temperaturas por debajo de 200 °C y mostrando excelente transparencia óptica. Aunque puede formarse con vacantes de oxígeno, estas se corrigen fácilmente con tratamientos de pasivación superficial. Su banda de conducción se

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

solapa de modo eficiente con la perovskita, reduciendo pérdidas por barrera de energía y mejorando la corriente de cortocircuito (J_{sc}).

ZnO (óxido de zinc): ZnO destaca por ofrecer una de las mayores movilidades de electrones reportadas para el ETL ($>100 \text{ cm}^2/\text{V}\cdot\text{s}$) y alta transmisión en el rango visible, lo que favorece la entrada de luz hacia la capa activa. No obstante, sin una pasivación adecuada puede reaccionar con la perovskita, generando sitios de recombinación en la interfaz. Su uso es habitual en dispositivos flexibles, pero en entornos húmedos requiere sellado o recubrimientos protectores.

PCBM ([6,6]-Fenil C_{61} ácido butírico metil ester): Este material orgánico permite procesar ETL a baja temperatura ($<100 \text{ }^\circ\text{C}$) y es especialmente valioso en sustratos flexibles o aplicaciones donde el calor es un limitante. Sin embargo, su robustez térmica y óptica es inferior a la de los óxidos inorgánicos, y suele requerir matrices de soporte para garantizar un transporte de carga estable a largo plazo.

Capa activa de perovskita

La perovskita compone la capa activa de la celda solar, la cual se basa en compuestos ABX_3 . Entre las perovskitas más usadas destacan el FAPbI_3 , MAPbI_3 y MASnI_3 , cuyas propiedades electrónicas y ópticas clave se recogen en la Tabla 2-2. El FAPbI_3 destaca por su banda prohibida de $\approx 1,48 \text{ eV}$ siendo una banda prohibida ideal donde se obtiene un balance óptimo entre corriente y voltaje. Además, un coeficiente de absorción superior a 10^5 cm^{-1} , que le permite capturar la mayor parte del espectro solar en apenas 300–500 nm de espesor [14]. Su fase α la llamada “perovskita negra” ofrece una estabilidad térmica sobresaliente, aunque exige un estricto control de la humedad durante el procesado. El plomo en su estructura reduce el número de defectos y favorece una fotoluminiscencia intensa, señal de una recombinación mínima de portadores. Por su parte, MAPbI_3 fue uno de los pioneros en perovskitas híbridas, alcanzando eficiencias cercanas al 20% con una síntesis sencilla y a temperatura moderada [17]. Sin embargo, su menor resistencia al calor y su sensibilidad a la humedad y a la irradiación lo hacen menos estable que FAPbI_3 , limitando su paso a aplicaciones de larga duración. Finalmente, MASnI_3 ofrece la gran ventaja de prescindir completamente del plomo y su banda prohibida ($\sim 1,30 \text{ eV}$) roza el óptimo teórico para maximizar la conversión fotovoltaica [18]. Pero la facilidad con la que el Sn^{2+} se oxida a Sn^{4+} provoca una mayor densidad de defectos y acorta la vida media de los portadores. Además, su procesado bajo atmósferas inertes y la necesidad de pasivaciones químicas complejas encarecen y dificultan su reproducibilidad en condiciones reales. Las propiedades electrónicas y ópticas clave de las perovskitas MAPbI_3 y MASnI_3 se recogen en la Tabla 2-2.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Tabla 2-2 Materiales de la capa activa

Característica	FAPbI ₃	MASnI ₃
Catión orgánico (A)	FA ⁺ (Formamidinio)	MA ⁺ (Metilamonio)
Catión metálico (B)	Pb ²⁺ (Plomo)	Sn ²⁺ (Estaño)
Haluro (X)	I ⁻ (Yoduro)	I ⁻ (Yoduro)
Eficiencia reportada	Alta (hasta ~25% en laboratorio)	Menor (~10 – 16%)
Estabilidad ambiental	Buena (mejor que MAPbI ₃)	Baja (Sn ²⁺ se oxida fácilmente)
Banda prohibida	~1.48eV	~1.3eV (más cercano al óptimo teórico)
Toxicidad	Alta (por plomo)	Menor (sin plomo)
Uso en dispositivos	Frecuente	Menos común
Procesabilidad	Compatible con técnicas húmedas	Similar, pero requiere atmosferas inertes
Coefficiente de absorción (α)	Alto (~10 ⁵ cm ⁻¹) desde 500–800 nm	Alto (~10 ⁵ cm ⁻¹), incluso en IR
Índice de refracción (n) en el visible	~ 2.3–2.6	~ 2.4–2.8
Fotoluminiscencia (PL)	Intensa, indica baja recombinación	Débil, más propenso a defectos
Tipo de portador mayoritario	Electrones (tipo n o ambipolar)	Huecos (típicamente tipo p)

Capa de transporte de huecos (HTL)

Spiro-OMeTAD: Se ha convertido en el estándar o el material más empleado en los laboratorios gracias a su excelente compatibilidad con perovskitas como FAPbI₃. Este material presenta una movilidad de huecos del orden de 10⁻⁴ cm²/V·s y una energía de nivel de valencia ajustada para minimizar la barrera de extracción. Su formulación, sin embargo, es relativamente costosa y requiere dopantes (LiTFSI, FK209) para alcanzar conductividades óptimas; además, muestra cierta sensibilidad a la humedad, lo cual requiere que las celdas que lo emplean estén encapsuladas para evitar la rápida degradación del mismo [19].

NiO_x (óxido de níquel): Como alternativa inorgánica aporta robustez ambiental y un comportamiento hidrófobo natural que mejora la durabilidad de las celdas solares. Tras dopados térmicos o químicos, sus vacantes de oxígeno generan huecos libres, alcanzando conductividades comparables a las de materiales orgánicos. Sin embargo, lograr uniformidad y baja resistividad requiere optimizar cuidadosamente el proceso de deposición y tratamiento, lo que puede introducir complejidad en la fabricación [20].

PEDOT:PSS (polietileno diacetato:poliestirensulfonato): Este polímero conductor es atractivo por su facilidad de procesamiento en ambiente y su bajo coste. Su energía de trabajo y su viscosidad permiten depositarlo por spin-coating sin necesidad de calor extremo. No obstante, su naturaleza higroscópica suele derivar en degradación bajo condiciones húmedas, razón por la que se combina con aditivos (GOPS, DMSO) o se recubre con capas protectoras para mejorar su estabilidad a largo plazo [21].

PTAA (politriaryl amina): Este material destaca por una movilidad de huecos superior a 10⁻³ cm²/V·s y por su notable resistencia frente a variaciones de temperatura y humedad. En este caso las películas son en general más uniformes, gracias a que es un material rígido. Sin

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

embargo, su coste y la menor difusión de protocolos en la literatura suelen relegarlo a aplicaciones más especializadas [22].

2.2 Parámetros eléctricos y ópticos en celdas solares

Para evaluar y optimizar una celda solar es imprescindible cuantificar tanto sus propiedades ópticas que determinan cuánta luz se aprovecha, así como sus parámetros eléctricos de desempeño que miden cómo esa luz se convierte en corriente y potencia.

Cuando un rayo de sol incide sobre la celda solar, solo los fotones con energía igual o superior a la banda prohibida (E_g) pueden crear pares electrón–hueco [23]. Esta banda prohibida está dada por la ecuación (2-1).

$$E_g = \frac{hc}{\lambda_g} \quad (2-1)$$

Donde λ_g es la longitud de onda de corte, h la constante de Planck y c la velocidad de la luz. Una E_g bien elegida ($\approx 1,3-1,5$ eV en perovskitas) asegura que se aproveche gran parte del espectro solar. De allí, que un factor importante en el diseño de este tipo de dispositivos es la elección de los materiales a emplear. Una vez superado ese umbral, la absorción de fotones por parte de la celda solar queda descrita por el coeficiente $\alpha(\lambda)$:

$$\alpha(\lambda) = A(h\nu - E_g)^m ; \text{ para } h\nu > E_g, \quad (2-2)$$

donde A es una constante empírica característica del material y m es un exponente que depende del tipo de transición óptica; en el caso de materiales con transición directa como las perovskitas utilizadas en este trabajo se asume $m=0.5$. Con ello, una vez ingresa en las diferentes capas que componen la celda solar, la intensidad de la radiación decrece por según la ley de Beer–Lambert:

$$I(d) = I_0 e^{-\alpha d}, \quad (2-3)$$

donde $I(d)$ representa la intensidad de la luz a una profundidad d dentro del material, mientras que I_0 es la intensidad inicial de la luz antes de ingresar al medio. El parámetro α corresponde al coeficiente de absorción del material, expresado en cm^{-1} , y d es el espesor o profundidad recorrida dentro del material, en centímetros. Esta relación permite calcular cómo disminuye la intensidad de la luz a medida que atraviesa un material absorbente [24]. Esto significa que a medida que la luz penetra en el material, su intensidad disminuye exponencialmente en función del espesor y del coeficiente de absorción. Además, cuanto mayor sea el coeficiente de absorción, más rápido se absorbe la luz, lo que es deseable en una celda solar de película delgada, esto es útil para estimar cuánta luz es absorbida en una capa determinada sin tener que resolver ecuaciones de Maxwell o considerar interferencias ópticas.

SCAPS-1D usa la ley Beer–Lambert para estimar cuánta energía luminosa es absorbida por cada capa del dispositivo, lo que a su vez se traduce en generación de portadores (electrones y huecos). Sin embargo, la ley no considera fenómenos como la reflexión, refracción ni

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

interferencias ópticas, por lo cual es una aproximación, lo cual es útil, pero limita su uso para estructuras ópticamente complejas [25].

Por otro lado, es importante tener en cuenta otros parámetros ópticos importantes asociados a las celdas solares de perovskita están asociados al índice de refracción $n(\lambda)$ y el coeficiente de extinción $k(\lambda)$, este último se relaciona directamente con la absorción de fotones del material como se muestra en la ecuación (2-5) [26].

$$n(\lambda) = \frac{c}{v} \quad (2-4)$$


$$k(\lambda) = \frac{\lambda\alpha}{4\pi} \quad (2-5)$$

En este sentido, controlar el índice de refracción en cada capa es esencial para diseñar recubrimientos anti-reflectantes, ya que al minimizar la reflectancia R y la transmitancia T , maximizamos la absorción A , tal que $A = 1 - R - T$. Así mismo, al controlar los valores de índice de refracción, se controla la dirección de los rayos solares y como estos se van refractando al interior de las diferentes capas que componen la celda solar.

Por otro lado, el comportamiento eléctrico de una celda solar describe cómo responde el dispositivo cuando es expuesto a la radiación solar y conectado a una carga externa, permitiendo la conversión de energía luminosa en energía eléctrica. Este comportamiento se representa comúnmente mediante la curva densidad de corriente-voltaje (J-V), la cual revela información clave sobre el desempeño del dispositivo bajo diferentes condiciones de operación. Cuando la luz incide sobre la celda, los fotones generan pares electrón-hueco que, al ser separados por el campo eléctrico interno de la unión $p-n$, producen una corriente eléctrica. La magnitud de esta corriente y el voltaje generado dependen de múltiples factores físicos, como la calidad de los materiales, las propiedades de las uniones, y las condiciones de iluminación. La caracterización eléctrica de una celda solar permite cuantificar su eficiencia, identificar pérdidas internas, y optimizar su diseño para maximizar la entrega de potencia a una carga, siendo por tanto esencial para comprender y mejorar el rendimiento global del dispositivo fotovoltaico [27].

Los siguientes parámetros eléctricos definen el comportamiento de una celda solar típica:

1. La corriente de cortocircuito (I_{sc}) es la máxima corriente generada por una celda solar cuando el voltaje entre sus terminales es cero, es decir, bajo condiciones de cortocircuito. Este parámetro está directamente relacionado con la cantidad de luz absorbida, ya que depende del número de fotones incidentes que logran generar pares electrón-hueco recolectables. En sistemas prácticos y simulaciones como las realizadas en SCAPS-1D, se emplea comúnmente la densidad de corriente de cortocircuito (J_{sc}), la cual se refiere a la corriente generada por unidad de área (A/cm^2) y permite una mejor comparación entre dispositivos de diferentes tamaños. La J_{sc} es una métrica clave en la caracterización de celdas solares, ya que refleja la eficiencia en la generación y recolección de carga a partir de la absorción de luz [28].

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

La expresión general para calcular la densidad de corriente de cortocircuito está dada por la ecuación (2-6). Además, la corriente de cortocircuito está dada por la ecuación (2-7), la cual integra la densidad de fotones que inciden sobre la celda, considerando el material y el espectro de luz incidente.

$$J_{sc} = q \int_{\lambda} \phi(\lambda) \cdot A(\lambda) \cdot \eta IQE(\lambda) d\lambda \quad (2-6)$$

$$I_{sc} = q \cdot \int_0^{\infty} \frac{E_{ph}(E)}{h \cdot c} \cdot \alpha dE, \quad (2-7)$$

donde q es la carga del electrón, $\phi(\lambda)$ representa el flujo de fotones incidentes por unidad de longitud de onda, $A(\lambda)$ es la fracción de luz absorbida por el dispositivo para cada longitud de onda considerada, $\eta IQE(\lambda)$ es la eficiencia cuántica interna, la cual indica qué proporción de los fotones absorbidos generan portadores de carga que efectivamente son recolectados, $E_{ph}(E)$ es la densidad espectral de flujo de fotones y E es la energía de los fotones [29].

- La tensión de circuito abierto (V_{oc}) representa el voltaje máximo que puede generar una celda solar cuando el circuito está completamente abierto, es decir, cuando no circula corriente. Este parámetro refleja el potencial eléctrico entre los contactos del dispositivo y está fuertemente influenciado por factores como la calidad del material, la alineación de bandas, la energía de banda prohibida (E_g) y el grado de recombinación de portadores (electrones y huecos). En celdas solares eficientes, un V_{oc} alto indica bajas pérdidas por recombinación, lo cual es especialmente deseable en tecnologías como las perovskitas o películas delgadas. En el entorno de simulación SCAPS-1D, V_{oc} se relaciona matemáticamente con la corriente de cortocircuito (I_{sc}), la corriente de saturación inversa (I_0), la temperatura (T), y constantes físicas fundamentales, según la siguiente expresión:

$$V_{oc} = \frac{K_B \cdot T}{q} \ln \left(\frac{I_{sc}}{I_0} + 1 \right) \quad (2-8)$$

donde K_B es la constante de Boltzmann y I_0 depende de las propiedades del material semiconductor. Esta relación permite estudiar cómo cambios en el diseño del dispositivo o en las condiciones de operación afectan el rendimiento fotovoltaico, particularmente en la optimización del V_{oc} [30].

- El punto de máxima potencia (P_{max}) es el punto en la curva I-V donde el producto de corriente (I) por voltaje (V) es máximo. Este punto indica la máxima potencia que una celda solar puede generar bajo condiciones específicas de iluminación. Esta P_{max} de acuerdo con la ley de Ohm, como se muestra en la ecuación (2-9).

$$P_{max} = V_{mp} \cdot I_{mp}, \quad (2-9)$$

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

donde V_{mp} e I_{mp} son respectivamente, el voltaje y la corriente en el punto de máxima potencia.

- El factor de llenado o fill factor (FF), es un parámetro clave en la evaluación del rendimiento de una celda solar, ya que mide la eficiencia con la que la energía capturada se convierte en potencia útil, el cual puede ser calculado empleando la ecuación (2-10). En términos prácticos, un FF elevado indica que la curva J-V del dispositivo se aproxima a una forma rectangular ideal, lo cual sugiere un comportamiento óptimo (véase Figura 2-2). Valores comerciales de FF suelen oscilar entre el 70% y el 85%, y su optimización es esencial para minimizar pérdidas internas y maximizar la eficiencia de conversión energética del dispositivo [31].

$$FF = \frac{P_{max}}{I_{sc} \cdot V_{oc}} \quad (2-10)$$

Existe una relación entre el FF y la resistencia en derivación (R_{sh} o R_{shunt}), la cual representa caminos alternativos para que la corriente se desvíe, como fugas por defectos o interfaces defectuosas. Una R_{sh} baja significa mayores pérdidas por fuga, lo que reduce la corriente disponible en el punto de máxima potencia, y por tanto disminuye el FF. No hay una única fórmula universal para expresar esta relación, porque el FF depende de múltiples factores no lineales, sin embargo, en ciertos modelos simplificados sí se puede expresar una relación aproximada que muestra cómo influye R_{sh} en el FF. En celdas con resistencia serie baja y bajo nivel de recombinación, el FF puede aproximarse mediante expresiones empíricas que incluyen R_s y R_{sh} . Una forma aproximada es:

$$FF \approx FF_{ideal} - \Delta FF_{RS} - \Delta FF_{RSh} \quad (2-11)$$

donde FF_{ideal} es el factor de llenado en condiciones ideales (sin pérdidas resistivas), el ΔFF_{RS} representa la pérdida por resistencia en serie, y el ΔFF_{Rsh} representa la pérdida por resistencia en derivación. La pérdida por R_{sh} puede expresarse como:

$$\Delta FF_{RSh} \approx \frac{V_{oc}}{I_{sc} - R_{sh}} \quad (2-12)$$

Esta ecuación demuestra que a menor R_{sh} , mayor será la pérdida de FF, especialmente cuando el voltaje de circuito abierto es alto [32].

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

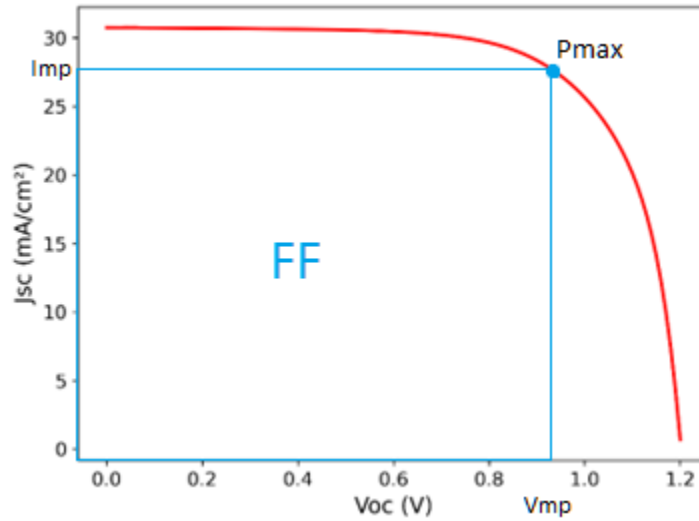



Figura 2-2. Curva J-V de una celda solar de perovskita, indicando como se determina el factor de llenado.

- La eficiencia de conversión de potencia (PCE, por sus siglas en inglés) es el parámetro más relevante en la evaluación de una celda solar, ya que representa el porcentaje de energía solar incidente que se convierte en energía eléctrica útil (P_{max}/P_{in}). Como se observa en la ecuación (2-10), la PCE puede expresarse como el producto de V_{oc} , J_{sc} y FF, dividido entre la potencia de entrada (P_{in}), bajo condiciones estándar (ecuación 2-13). La PCE es una medida global del desempeño de la celda, ya que resume la interacción entre los procesos ópticos y eléctricos. La mejora de la PCE implica una optimización simultánea de los tres parámetros anteriores, por lo cual se considera el objetivo final en el desarrollo de cualquier tecnología fotovoltaica. No obstante, la PCE también está limitada por consideraciones termodinámicas (Shockley-Queisser), por lo que alcanzar altas eficiencias requiere estrategias avanzadas, como estructuras tándem, texturizado óptico o pasivación superficial [33].

$$PCE = \frac{P_{max}}{P_{in}} \times \frac{V_{oc} \cdot J_{sc} \cdot FF}{P_{in}} \quad (2-13)$$

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

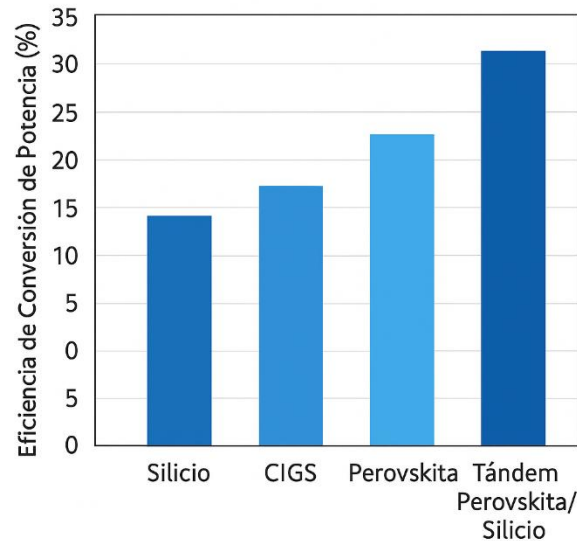


Figura 2-3. Comparación de eficiencia en diferentes tecnologías solares.

Las celdas solares de tipo tándem, especialmente las que combinan perovskita con silicio, han alcanzado las mayores PCE registradas en laboratorio. Estas celdas logran una eficiencia mayor al 34.6%, como ha sido reportado actualmente en 2025, esta alta eficiencia se debe a la capacidad de la celda de tándem de captar diferentes longitudes de onda de la luz solar utilizando dos materiales complementarios [1]. En segundo lugar, se encuentran las celdas solares de perovskita de monocapa, las cuales han logrado una eficiencia máxima de 26.95% [1]. Las celdas solares de silicio cristalino avanzado, como las tecnologías tipo TOPCon, IBC (Interdigitated Back Contact) o HJT (Heterojunction with Intrinsic Thin layer), alcanzan eficiencias de hasta 27.6%, según informes que la gráfica del NREL reporta actualmente [1]. Finalmente, las celdas solares CIGS (cobre, indio, galio y selenio), que son una tecnología de película delgada, han alcanzado eficiencias de laboratorio cercanas al 23.64%, según registros históricos certificados por Fraunhofer ISE [34]. Aunque son menos eficientes, ofrecen ventajas en flexibilidad y producción en sustratos livianos. Esta información está respaldada por la gráfica de eficiencias del NREL, ampliamente usada en la comunidad científica para visualizar los récords de conversión fotovoltaica por tecnología.

Impacto del plomo y retos de la perovskita de estaño

Las perovskitas de plomo (Pb^{2+}) alcanzan eficiencias superiores al 25%, pero la toxicidad del Pb^{2+} supone un serio riesgo ambiental y sanitario tras posibles rupturas o incorrecto reciclaje [35]. La encapsulación y los programas de recuperación minimizan la liberación de plomo, pero elevan notablemente los costes y complican la línea de producción.

El estaño (Sn^{2+}) aparece como sustituto no tóxico, con propiedades ópticas y eléctricas comparables. No obstante, su rápido paso a Sn^{4+} introduce defectos profundos en la estructura cristalina que actúan como centros de recombinación no radiactiva, reduciendo drásticamente

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

la eficiencia [36]. Además, las películas de Sn-perovskita demandan atmósferas inertes y aditivos antioxidantes para evitar oxidación, lo que limita el escalado industrial.

Por si fuera poco, estas perovskitas presentan mayor densidad de vacancias y una sensibilidad extrema a la humedad y al oxígeno, acelerando la degradación bajo condiciones de operación reales [36]. Superar estos desafíos, es decir, mejorar la estabilidad química, reducir defectos y simplificar el procesamiento es crucial para que las perovskitas de estaño resulten una alternativa sostenible y comercialmente viable.

2.3 Herramientas de simulación para dispositivos fotovoltaicos.

Las herramientas de simulación para dispositivos fotovoltaicos desempeñan un papel fundamental en la investigación, diseño y optimización de celdas solares, permitiendo predecir y analizar el comportamiento eléctrico, óptico y estructural de estos dispositivos antes de su fabricación física. Aunque estos simuladores como SCAPS-1D, TCAD, COMSOL o Silvaco no ofrecen resultados absolutamente precisos debido a simplificaciones numéricas, supuestos ideales, limitaciones dimensionales (como en modelos 1D) y falta de parámetros experimentales exactos, siguen siendo valiosos por su capacidad de reducir costos y tiempo en el desarrollo de nuevas tecnologías solares. Estas herramientas permiten visualizar el impacto de parámetros como el dopaje, los espesores de capas, las barreras de potencial, las tasas de recombinación, y las propiedades de transporte de carga, lo que facilita la identificación de cuellos de botella y la exploración de configuraciones óptimas sin necesidad de fabricar prototipos reales en cada iteración. Además, su accesibilidad ha permitido una expansión significativa del estudio fotovoltaico en instituciones académicas de todo el mundo, incluso en entornos con recursos limitados. A continuación, se presenta una breve revisión de las principales herramientas computacionales que se emplean en la actualidad para el modelamiento de estos componentes:

TCAD (Technology Computer-Aided Design)

Es una herramienta de simulación avanzada, utilizada para modelar dispositivos semiconductores desde un enfoque físico riguroso. Permite realizar análisis eléctricos, ópticos y térmicos en configuraciones unidimensionales, bidimensionales y tridimensionales, siendo particularmente útil para tecnologías como celdas solares de silicio, perovskita, tandem y otras arquitecturas emergentes. Este software requiere la definición explícita de modelos físicos como recombinación, generación y transporte de carga, lo que le otorga un alto grado de precisión y flexibilidad. No obstante, esta capacidad también implica un considerable consumo computacional, ya que las simulaciones —especialmente en 3D o en régimen transitorio— pueden tardar desde varias horas hasta días, dependiendo de la complejidad del dispositivo y de los parámetros modelados. Aunque TCAD destaca por su capacidad de ofrecer resultados detallados y altamente personalizables, presenta limitaciones importantes como una curva de aprendizaje pronunciada y un costo elevado de la licencia [37].

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

COMSOL Multiphysics


Este es un software de simulación numérica de propósito general diseñado para resolver sistemas de ecuaciones físicas en geometrías complejas. Su estructura modular permite integrar distintos fenómenos físicos como transferencia de calor, óptica, mecánica estructural y transporte de carga eléctrica, lo que lo hace especialmente útil en el estudio de celdas solares cuando se requiere un enfoque multiprofísico, por ejemplo, para analizar simultáneamente los efectos térmicos, ópticos y eléctricos. Dispone de una interfaz gráfica intuitiva que facilita la construcción de modelos, aunque también permite programación avanzada mediante scripting en MATLAB. En términos de demanda computacional, COMSOL puede requerir una gran cantidad de recursos, especialmente cuando se trabaja con modelos tridimensionales o con múltiples ecuaciones acopladas. Sus principales ventajas incluyen su flexibilidad y facilidad para abordar problemas acoplados, aunque su uso efectivo exige conocimientos sólidos en formulación matemática de fenómenos físicos. Además, su esquema de licencias especializadas por módulos incrementa significativamente el costo total del software [37], [38].

Silvaco ATLAS

Silvaco ATLAS es una herramienta de simulación perteneciente a la suite TCAD de Silvaco, enfocada en el modelado de dispositivos electrónicos y optoelectrónicos con un enfoque físico riguroso. Ofrece capacidades similares a TCAD Synopsys, pero se destaca por una interfaz relativamente más accesible para usuarios con conocimientos intermedios, facilitando la exploración de tecnologías avanzadas como celdas solares de perovskita, CIGS o estructuras tándem. El software permite simular materiales semiconductores novedosos, así como incorporar efectos físicos complejos como recombinación Shockley-Read-Hall, efecto fotovoltaico, transporte dependiente del campo y fenómenos cuánticos. En términos de rendimiento, presenta un consumo computacional de moderado a alto, dependiendo de la complejidad del modelo y la dimensionalidad: las simulaciones típicas en 2D pueden requerir varias horas, mientras que en 3D demandan recursos computacionales significativamente mayores. Su principal ventaja es el equilibrio entre profundidad física y facilidad de uso, lo que ha llevado a su adopción en numerosos proyectos de investigación aplicada. No obstante, se debe considerar que su uso está restringido por licencias comerciales costosas y exige un nivel técnico considerable para configurar modelos con precisión [37].

SCAPS-1D (Solar Cell Capacitance Simulator)

Esta herramienta es un software especializado en la simulación de dispositivos fotovoltaicos de estructura unidimensional, ampliamente utilizado en el ámbito académico para el análisis de tecnologías como las celdas solares de perovskita, CdTe, CIGS y otros dispositivos de película delgada. Su enfoque está centrado en la simulación de parámetros eléctricos clave como curvas densidad de corriente-voltaje (J-V), capacitancia, perfiles de potencial eléctrico y distribución de portadores, lo que lo convierte en una herramienta efectiva para estudios

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

paramétricos y comparativos. Destaca por su interfaz gráfica sencilla y su accesibilidad, ya que es gratuito para fines académicos. Una de sus principales ventajas es su bajo consumo computacional: puede ejecutarse en prácticamente cualquier computador con especificaciones modestas (2–4 GB de RAM), y las simulaciones suelen completarse en segundos o minutos, permitiendo realizar múltiples barridos de parámetros con alta eficiencia temporal. No obstante, su simplicidad también implica limitaciones importantes: sólo permite simulaciones en una dimensión (1D), y no contempla efectos térmicos ni ópticos avanzados, lo que restringe su aplicabilidad en análisis más complejos o realistas. A pesar de ello, su facilidad de uso y eficiencia lo han consolidado como una herramienta clave en investigaciones académicas, especialmente en etapas preliminares de diseño o validación conceptual de celdas solares [39].

Tabla 2-3 Comparación general de herramientas de simulación

Software	Dimensiones	Precisión	Complejidad	Tipo de licencia
TCAD	1D/2D/3D	Muy alta	Alta	Comercial
COMSOL	1D/2D/3D	Muy alta	Muy alta	Comercial
Silvaco ATLAS	1D/2D/3D	Alta	Media-Alta	Comercial
SCAPS-1D	1D	Media	Muy baja	Gratuito


Debido a que, en esta investigación, se requiere realizar un alto volumen de simulaciones en un tiempo mínimo, SCAPS-1D representa una de las opciones más eficientes desde el punto de vista computacional, debido a su arquitectura simplificada basada en modelos unidimensionales y su bajo requerimiento de recursos de hardware.

2.4 Aprendizaje de máquina en la optimización de celdas solares.

En los últimos años, el aprendizaje de máquina se ha consolidado como una herramienta poderosa para abordar problemas de optimización en distintos campos de la ingeniería [40], [41], [42], [43], [44], [45]. En este contexto de las celdas solares, su aplicación permite modelar y predecir el comportamiento de dispositivos fotovoltaicos a partir de grandes volúmenes de datos experimentales o simulados, facilitando el diseño eficiente y la mejora del rendimiento.

Revisión general de tareas de regresión en sistemas físicos complejos.

Las tareas de regresión, dentro del ámbito del aprendizaje de máquina, se utilizan para estimar variables continuas en función de múltiples características de entrada. En sistemas físicos complejos como las celdas solares, donde múltiples parámetros interactúan de forma no lineal, los algoritmos de regresión permiten capturar estas relaciones y realizar predicciones precisas sobre métricas clave como la eficiencia de conversión o el factor de llenado. Esta sección presenta una visión general de las principales técnicas empleadas y su aplicabilidad en el estudio y optimización de estos dispositivos.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Diversos autores han abordado el problema de optimización de celdas solares mediante técnicas de regresión en sistemas físicos complejos, utilizando modelos de aprendizaje de máquina aplicados a datos simulados. Por ejemplo, Shrivastav y colaboradores optimizaron el desempeño de celdas solares de perovskita basadas en cesio (Cs-PSCs) mediante simulaciones en SCAPS-1D y ML [46]. En este trabajo los autores analizaron seis materiales de capa absorbente en función del espesor de la capa absorbente, la densidad de dopaje y la densidad de defectos. Se entrenaron cinco modelos de ML diferentes Regresión Lineal, Máquinas de Vectores de Soporte (Support Vector Regression, SVR), Redes Neuronales (Neural Networks, NN), Random Forest y Extreme Gradient Boosting (XGBoost). Siendo XGBoost el de mejor desempeño ($R^2 = 99.99\%$, $MSE = 0.0006$). Los resultados revelaron que el material y el espesor de la capa absorbente fueron los factores más influyentes en la eficiencia, mientras que la densidad de dopaje tuvo menor impacto. De forma análoga, Jiang y compañía emplearon ML para optimizar el dopaje de yoduro de potasio (KI) en celdas solares MAPbI₃, evitando métodos de prueba y error [47]. Se encontró que una concentración ideal del 3% de KI disminuye la cantidad de defectos, logrando una eficiencia máxima del 20.91%. Esto muestra cómo el ML puede ayudar a mejorar más rápidamente los materiales fotovoltaicos. En 2025, Deswal y su equipo mejoraron una celda solar de perovskita doble sin plomo, que usa Cs₂RbGaI₆, a través de simulaciones en SCAPS-1D [48]. Se calcularon propiedades clave como la densidad de estados y la movilidad de portadores utilizando Teoría del Funcional de la Densidad (DFT, por sus siglas en inglés), mientras que la afinidad electrónica, densidad de defectos y espesor de la capa absorbente fueron optimizados en SCAPS-1D. Se desarrolló un modelo de regresión de bosques aleatorios con 1008 datos simulados, logrando un R^2 de 0.9985, y se utilizaron los resultados para analizar la influencia del espesor y defectos en la PCE, que alcanzó un máximo de 22.17% [48].

Basado en lo anterior se, demuestran que los modelos de ML pueden capturar relaciones no lineales entre variables estructurales y ópticas en dispositivos fotovoltaicos, permitiendo predicciones precisas de la eficiencia sin necesidad de realizar ensayos experimentales exhaustivos.

Explicación y estructura del modelo: arboles de Clasificación y Regresión (CART)

Los árboles de decisión, especialmente los que se construyen con el algoritmo CART, son modelos que no necesitan asumir una forma específica en los datos. Esto los hace muy útiles para descubrir relaciones complejas entre varias variables que usamos para predecir, y la variable que queremos predecir. Estos árboles funcionan dividiendo los datos en dos partes cada vez (divisiones binarias) [49], de forma que los grupos que se forman sean lo más parecidos posible en cuanto al valor que queremos predecir.

Cuando el objetivo es predecir un número (regresión), el árbol busca en cada paso la mejor variable y el mejor punto de corte que hagan que los errores entre los valores reales y los predichos sean lo más pequeños posible. Esta diferencia se mide usando una fórmula llamada

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

RSS (suma de los errores al cuadrado). Por otro lado, cuando el objetivo es clasificar (por ejemplo, decidir si algo es de una clase u otra), el árbol usa otras medidas, como el índice Gini, la entropía o el error de clasificación, los cuales ayudan a elegir las divisiones que hacen que los grupos sean lo más puros posible (es decir, que contengan elementos de una sola clase) [50]. El criterio de reducción de impureza para clasificación está dado por:

$$I = I_a - (\pi_1 I_1 + \pi_2 I_2) \quad (2-14)$$

Donde I_a es la impureza del nodo antes de la partición, I_1 e I_2 son las impurezas de los nodos hijos, y π_1 y π_2 representan las proporciones de observaciones en cada nodo. Este proceso se repite hasta que se cumple un criterio de parada, como un número mínimo de observaciones por nodo o una reducción de impureza insignificante.

Una vez construido el árbol, las predicciones se hacen a partir del promedio (en regresión) o de la clase mayoritaria (en clasificación) de las observaciones contenidas en cada nodo terminal. Los árboles de decisión destacan por su capacidad de modelar relaciones no lineales y efectos de interacción sin requerir especificación previa, lo que los convierte en una herramienta flexible. Además, pueden manejar variables categóricas sin transformación y son relativamente robustos a valores faltantes mediante divisiones sustitutas.

Desventajas de los árboles de decisión

A pesar de su simplicidad y capacidad de interpretación, los árboles de decisión presentan limitaciones importantes que restringen su desempeño en tareas predictivas complejas. Una de las principales desventajas es su alta varianza. Dado que los árboles se construyen a partir de divisiones locales altamente dependientes de los datos de entrenamiento. Como consecuencia pequeñas perturbaciones en estos pueden resultar en árboles muy diferentes, afectando la estabilidad del modelo y su capacidad de generalización [51]. Este comportamiento errático compromete no solo la precisión de las predicciones, sino también la interpretabilidad de los resultados, ya que no se garantiza una estructura consistente del árbol entre diferentes muestras.

Para mitigar esta inestabilidad, una estrategia clásica es la poda, que consiste en construir primero un árbol profundo y luego reducir su tamaño eliminando nodos internos poco informativos. Este procedimiento, conocido como *cost-complexity pruning*, busca un equilibrio entre el ajuste al conjunto de entrenamiento y la capacidad de generalizar a nuevos datos [52].

Otro problema estructural de los árboles CART es su sesgo en la selección de variables. Este sesgo favorece a aquellas variables que tienen mayor número de categorías o menor cantidad de valores faltantes, ya que tienden a ofrecer más posibles divisiones y, por tanto, mayores reducciones de impureza [53]. En aplicaciones prácticas, este fenómeno puede conducir a una sobreestimación del poder predictivo de variables categóricas con muchos niveles o numéricas con pocas observaciones perdidas.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Finalmente, los árboles de decisión tienen dificultades para capturar estructuras aditivas entre variables predictoras, es decir, relaciones que pueden modelarse como la suma de efectos individuales. Dado que la construcción del árbol se basa en divisiones binarias sucesivas, esta arquitectura no se ajusta naturalmente a modelos donde los efectos son acumulativos y suaves. En estos casos, los métodos basados en regresión lineal o aditiva pueden ser más apropiados [54].

Bosques aleatorios: fundamentos y construcción

Los bosques aleatorios fueron introducidos por Breiman en el 2001 como una extensión del método *Bagging* (Bootstrap Aggregation), con el objetivo de reducir la varianza de los árboles individuales sin incrementar su sesgo. Al tratarse de un método de ensamble, Random Forest combina múltiples árboles de decisión, entrenados sobre subconjuntos aleatorios de los datos, y promedia sus resultados para obtener una predicción más robusta y estable.

En regresión, el modelo final corresponde al promedio de las predicciones de todos los árboles, mientras que en clasificación se selecciona la clase más votada. Este esquema permite capturar relaciones no lineales y complejas entre las variables predictoras y la respuesta, al mismo tiempo que mejora la capacidad de generalización del modelo. Una ventaja adicional es su capacidad para manejar datasets con un gran número de variables, incluidas aquellas con correlaciones internas, y su flexibilidad para calcular medidas de importancia, lo cual es esencial en contextos como el de esta tesis, donde se busca identificar los factores más relevantes en la eficiencia de celdas solares.

Aleatorización: doble fuente de variabilidad

El poder predictivo y la estabilidad de Random Forest provienen de dos mecanismos clave de aleatorización:

Remuestreo de los datos (bootstrap): cada árbol se entrena con una muestra aleatoria con reemplazo del conjunto de entrenamiento. Esto significa que ciertos registros pueden repetirse, mientras que otros quedan excluidos. En promedio, alrededor del 63.2% de las observaciones son incluidas en cada muestra bootstrap, dejando el restante 36.8% como conjunto de prueba interno, conocido como observaciones *out-of-bag* (OOB).

Selección aleatoria de variables (m_{try}): en cada nodo del árbol, en lugar de considerar todas las variables disponibles para realizar la partición, se selecciona aleatoriamente un subconjunto de tamaño m_{try} . Este valor suele fijarse en $\frac{p}{3}$ para regresión y en \sqrt{p} para clasificación, donde p es el número total de variables predictoras. Este procedimiento disminuye la correlación entre los árboles del conjunto, y con ello reduce la varianza del modelo ensamble, sin afectar negativamente el sesgo. Es importante resaltar que el error de generalización de un Random Forest depende de la precisión de los árboles individuales y del grado de correlación entre ellos [55].

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Relación entre varianza y correlación

La reducción de varianza en Random Forest puede expresarse mediante la siguiente relación:

$$V[\bar{X}] = \rho\sigma^2 + \frac{1-\rho}{\eta}\sigma^2, \quad (2-15)$$

donde σ^2 es la varianza de cada árbol, η es el número de árboles en el bosque, y ρ representa la correlación promedio entre los árboles. Esta ecuación muestra que, al disminuir ρ , se reduce la varianza total del modelo, lo cual es una de las razones fundamentales para aplicar la aleatorización en la selección de variables en cada nodo.

Construcción del bosque

El proceso de entrenamiento de un Random Forest puede resumirse en los siguientes pasos:

1. Para cada árbol $t \in \{1, \dots, n_{tree}\}$ seleccionar una muestra bootstrap del conjunto de entrenamiento.
2. Construir un árbol completo (sin poda) usando únicamente la muestra seleccionada.
3. En cada nodo, seleccionar aleatoriamente “*mtry*” variables y elegir aquella que genere la mayor ganancia (reducción del RSS en regresión o de impureza en clasificación).
4. Repetir este procedimiento para los n_{tree} árboles del bosque.

Este esquema asegura diversidad entre los árboles, permitiendo que el conjunto capture múltiples perspectivas del espacio de datos. Además, al entrenar cada árbol sobre una muestra distinta, se obtiene una estimación automática del error generalizado mediante las observaciones OOB, sin necesidad de una validación cruzada externa.

Los bosques aleatorios permiten estimar de forma automática la importancia de las variables utilizadas en el modelo. Esta estimación puede realizarse a través del índice de Gini o mediante la técnica de permutación aleatoria de las variables [55]. Ambas estrategias tienen como objetivo cuantificar la contribución de cada variable predictora sobre la respuesta, lo que permite identificar su relevancia y capacidad explicativa. Esta información resulta fundamental para el desarrollo de algoritmos de selección de variables, como el que se presenta a continuación.

Medida de importancia por permutación

Una de las formas más utilizadas para estimar la importancia de una variable en Random Forest consiste en evaluar cuánto se incrementa el error de predicción cuando se rompe su asociación con la variable respuesta (Y). Este procedimiento se realiza mediante permutación aleatoria [56].

Selección de variables con Random Forest

El algoritmo Random Forest permite incorporar de manera interna mecanismos de selección de variables durante el proceso de entrenamiento. A través de medidas como la importancia por permutación y la reducción del índice de Gini, el modelo es capaz de cuantificar el aporte

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

de cada predictor a la respuesta, permitiendo así identificar variables relevantes sin necesidad de aplicar técnicas de selección externas [50], [55].

Este enfoque, conocido como método embebido, ofrece una ventaja clave debido a que la selección de variables ocurre de forma simultánea al ajuste del modelo, considerando las interacciones entre predictores y su efecto conjunto sobre la variable objetivo. En el caso particular de este trabajo, donde se evalúa el rendimiento de celdas solares en función de parámetros físicos como los espesores de distintas capas, esta capacidad resulta esencial para detectar características redundantes o irrelevantes que puedan afectar la precisión del modelo.

Impacto de la correlación.

El estudio de Gregorutti [57] demuestra que la presencia de variables predictoras altamente correlacionadas puede distorsionar significativamente las medidas de importancia por permutación generadas por Random Forest. En contextos donde varias variables comparten información redundante, la permutación de una de ellas no genera un aumento sustancial en el error de predicción, dado que las demás pueden compensar su ausencia. Como resultado, estas variables correlacionadas tienden a recibir valores de importancia artificialmente bajos, incluso si son altamente relevantes para la variable objetivo. Este fenómeno, conocido como “sesgo de correlación”, se intensifica cuando aumenta el número de predictores correlacionados y puede llevar a que variables independientes, pero menos informativas, aparezcan como más importantes. A través de un análisis teórico en un modelo de regresión aditivo, los autores evidencian que este sesgo es sistemático y proponen como alternativa el uso de algoritmos de eliminación recursiva de características (RFE), que actualizan las medidas de importancia en cada iteración y permiten una selección de variables más robusta frente a la correlación.

Entrenamiento del modelo: RandomForestRegressor

El estimador *RandomForestRegressor* implementa el algoritmo de bosques aleatorios para resolver problemas de regresión. Su funcionamiento se basa en entrenar un conjunto de árboles de decisión sobre diferentes subconjuntos de los datos, generados mediante muestreo aleatorio con reemplazo (bootstrap). Cada árbol del conjunto aprende una función de predicción a partir de un subconjunto de observaciones y, en cada nodo, considera solo una fracción aleatoria de las variables disponibles, lo que introduce diversidad en la construcción del bosque.

Durante el entrenamiento, cada árbol T_t se ajusta para minimizar el error cuadrático en su muestra bootstrap, dividiendo los datos de manera recursiva según el valor de las variables predictoras. La predicción final del modelo es obtenida como el promedio de las predicciones individuales de todos los árboles:

$$\bar{y}_i^{(t)} = \frac{1}{n_{tree}} \sum_{t=1}^{n_{tree}} \bar{y}_i^{(t)} \quad (2-16)$$

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Donde $\hat{y}_i^{(t)}$ representa la predicción del árbol t para la observación i .

El proceso de entrenamiento incluye el cálculo del error OOB, estimado a partir de las observaciones no utilizadas en la muestra Bootstrap de cada árbol. Este error proporciona una estimación interna del rendimiento general del modelo sin requerir un conjunto de validación adicional. Paralelamente, el modelo calcula medidas de importancia de variables, como la importancia por permutación o la reducción del índice Gini, las cuales permiten identificar qué predictores tienen mayor influencia sobre la variable de salida.

Métricas de desempeño

Una vez entrenado el modelo, es necesario evaluar su capacidad predictiva mediante métricas de desempeño, las cuales se conocen en el ámbito de modelos de optimización y aprendizaje automático porque son funciones o medidas cuantitativas que permiten evaluar el rendimiento o calidad de un modelo. Una métrica es una herramienta que se utiliza para cuantificar qué tan bien está funcionando un modelo respecto a un objetivo específico, como la precisión, el error o la eficiencia. En otras palabras, estas métricas permiten cuantificar qué tan cerca están las predicciones del modelo respecto a los valores reales de la variable objetivo. En problemas de regresión como el abordado en este trabajo, las medidas más utilizadas son el coeficiente de determinación (R^2), el error cuadrático medio (MSE) y la raíz del error cuadrático medio (RMSE).

Coeficiente de determinación (R^2)

El coeficiente de determinación, conocido como R^2 , es una métrica que indica qué proporción de la varianza total de la variable dependiente puede ser explicada por el modelo. Varía entre 0 y 1, donde 1 representa una predicción perfecta y 0 indica que el modelo no explica nada mejor que la media de los datos. Un R^2 alto sugiere que el modelo captura la relación entre las variables independientes y la variable objetivo, mientras que un valor bajo indica que gran parte de la variabilidad no está siendo explicada [58].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2-17)$$

donde y_i es el valor real, \hat{y}_i es el valor predicho, \bar{y} la media de los valores reales y n el número de datos.

Error Absoluto Medio (MAE)

El MAE (mean absolute error) es una métrica que mide el promedio de los errores absolutos entre los valores predichos y los valores reales. Es decir, toma la diferencia entre ambos, ignora su signo y calcula el promedio total. Esta métrica proporciona una idea directa del error promedio que está cometiendo el modelo, expresado en las mismas unidades que la variable de salida. Su interpretación es sencilla: un MAE de 2, por ejemplo, indica que, en

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

promedio, el modelo se desvía 2 unidades de la realidad. Esta métrica es menos sensible a valores atípicos que otras, por lo que es útil en casos donde los errores extremos no deben influir demasiado en la evaluación [59]. Matemáticamente esta métrica puede ser estimada con la siguiente ecuación:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2-18)$$

Raíz del error cuadrático medio (RMSE)

Es una métrica que calcula la raíz cuadrada del promedio de los errores al cuadrado. A diferencia del MAE, penaliza con más severidad los errores grandes, ya que al elevar al cuadrado las diferencias, cualquier valor atípico tendrá un peso mucho mayor. Esta propiedad lo convierte en un excelente indicador cuando los errores grandes son especialmente costosos o críticos. La importancia del RMSE radica en su capacidad para detectar variaciones significativas entre el modelo y la realidad, ayudando a refinar modelos que deben ser extremadamente precisos. Además, es muy utilizada por su capacidad de traducir el error a la misma escala de la variable objetivo, facilitando su interpretación. Al usar RMSE, se puede evaluar si el modelo comete errores importantes y, en consecuencia, si es necesario aplicar mejoras como ajuste de hiperparámetros o limpieza de datos [60]. Matemáticamente esta métrica puede ser estimada con la siguiente ecuación:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2-19)$$

Optimización mediante hiperparámetros

La optimización de hiperparámetros es un paso fundamental en el entrenamiento de modelos de aprendizaje automático para mejorar su desempeño predictivo y garantizar una generalización adecuada. Los hiperparámetros son configuraciones externas al modelo, tales como la tasa de aprendizaje, la profundidad máxima de un árbol o el número de estimadores en un bosque aleatorio, que deben ser definidos antes del proceso de entrenamiento. El ajuste adecuado de estos hiperparámetros influye directamente en la capacidad del modelo para capturar patrones relevantes y evitar problemas de sobreajuste o subajuste.

El objetivo de la optimización es encontrar el conjunto de hiperparámetros (λ^*) que maximice una función objetivo, la cual puede ser una métrica de evaluación como la precisión, exactitud, o el error cuadrático medio, entre otras.

El método de búsqueda en malla (*grid search*) es una técnica sistemática que consiste en evaluar exhaustivamente el rendimiento del modelo en un conjunto discreto y predefinido de combinaciones de hiperparámetros. La evaluación del desempeño se realiza generalmente mediante técnicas de validación cruzada, en la cual el conjunto de datos se divide para medir la estabilidad y generalización del modelo.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

2.5 Técnicas de optimización multiobjetivo


La optimización multiobjetivo es una estrategia utilizada cuando se necesita mejorar varias metas al mismo tiempo, pero estas metas pueden entrar en conflicto entre sí. Por ejemplo, se podría querer maximizar el rendimiento de un sistema y al mismo tiempo minimizar su costo, lo que implica tomar decisiones equilibradas porque mejorar uno puede empeorar el otro. En lugar de buscar una única mejor solución, esta técnica permite encontrar un conjunto de soluciones óptimas que representan diferentes formas de alcanzar un buen equilibrio. Esto es especialmente útil en problemas complejos del mundo real, donde rara vez existe una única respuesta ideal. Estas soluciones forman lo que se conoce como un frente de Pareto, que muestra todas las alternativas igualmente válidas según los criterios definidos [61].

En los procesos de toma de decisiones complejas, donde intervienen múltiples intereses o criterios, rara vez existe una única solución que satisfaga completamente todas las necesidades. Por lo general, mejorar un aspecto puede implicar que otro se vea afectado. En estos contextos, se recurre a herramientas y enfoques de optimización que ayudan a visualizar las posibles soluciones y elegir entre ellas según prioridades específicas. Dos conceptos ampliamente utilizados en este tipo de problemas son el frente de Pareto y la maximización ponderada, que permiten explorar y comparar opciones viables de manera estructurada y razonada.

Concepto de Frente de Pareto.

El frente de Pareto representa un conjunto de soluciones que ofrecen los mejores compromisos posibles entre dos o más objetivos que compiten entre sí. En otras palabras, una solución está en este frente si no se puede mejorar en un criterio sin empeorar en otro.

La figura 2-6 representa un frente de Pareto en un problema de optimización con dos objetivos que se desean minimizar simultáneamente, F1 y F2. Cada punto en el gráfico corresponde a una solución posible, y la curva azul traza las soluciones llamadas no dominadas, es decir, aquellas para las que no existe otra solución que sea mejor en ambos objetivos al mismo tiempo. Por ejemplo, si una solución tiene un valor más bajo en F1, pero uno más alto en F2 comparada con otra, no puede decirse que una domine completamente a la otra; ambas están en compromiso. Sin embargo, si una solución es mejor o igual en todos los objetivos y estrictamente mejor en al menos uno, entonces domina a las demás. Esas soluciones dominadas se representan con círculos vacíos, mientras que las no dominadas (las mejores en equilibrio) se muestran con círculos rellenos y forman lo que se conoce como la frontera de Pareto. Esta visión permite seleccionar la solución más conveniente según la necesidad específica, sin imponer una sola respuesta ideal [62].

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

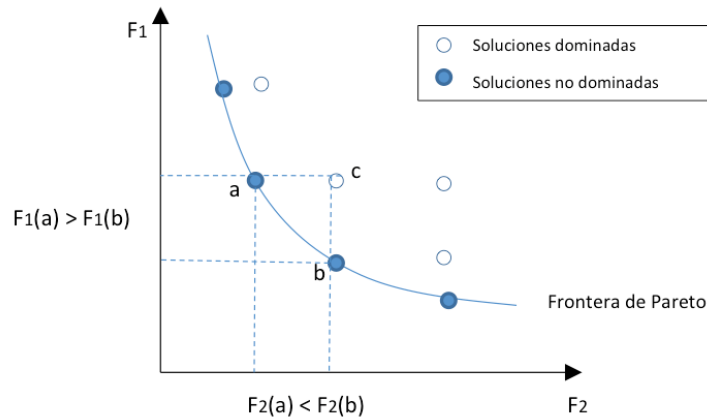


Figura 2-6 Frente de Pareto ideal en un problema de minimización [63].

En el caso de este trabajo, en donde se pretende diseñar una celda solar, podríamos querer maximizar tanto el FF, que mide la calidad interna de la celda, como la PCE. Sin embargo, a medida que se hacen ajustes para aumentar el FF como mejorar los contactos o reducir pérdidas internas podría que esos cambios no favorezcan del todo la absorción de luz o la recolección de carga, lo que afecta la PCE. Así, las soluciones óptimas no son únicas, por lo que el frente de Pareto agrupa las combinaciones de FF y PCE donde no se puede aumentar uno sin perder rendimiento en el otro. Como se maximiza FF y PCE, el frente de Pareto estaría compuesto por las combinaciones que tienen los valores más altos posibles de ambas métricas sin que una mejore a costa de empeorar la otra.

Maximización ponderada.

La maximización ponderada es una técnica común en problemas de optimización multicriterio, la cual permite combinar varios objetivos en una única expresión matemática. Cada objetivo se representa como una función y se le asigna un peso que refleja su importancia relativa dentro del proceso de decisión. Formalmente, esta combinación se define como:


$$F(x) = w_1 f_1(x) + w_2 f_2(x) + \dots w_k f_k(x) \quad (2-20)$$

Donde x representa el vector de parámetros de entrada (por ejemplo, los espesores de capas de una celda solar), $f_2(x)$ son las funciones objetivo y w_1 son los pesos asignados a cada una, sujetos a la condición $\sum_{i=1}^k w_i = 1$, donde $w_i \geq 0$.

En el contexto de este trabajo, si bien la eficiencia de conversión (PCE) es el objetivo principal, también se valora el factor de llenado (FF) como indicador de estabilidad del dispositivo. Por tanto, se puede construir una función objetivo ponderada donde, por ejemplo, se asigne un 70% de peso al FF y un 30% al PCE. El algoritmo de entrenamiento busca entonces maximizar esa combinación equilibrada, guiado por las prioridades definidas por los pesos.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Este método resulta útil cuando se busca automatizar la toma de decisiones considerando múltiples criterios, ya que permite establecer un compromiso explícito entre ellos. Sin embargo, también implica que los pesos asignados deben elegirse cuidadosamente, ya que influyen directamente en la solución final [64]. Una de las ventajas de este método es su flexibilidad: los pesos pueden ajustarse dinámicamente según las necesidades del sistema, el contexto de aplicación o las preferencias del diseñador.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

3. METODOLOGÍA

3.1 Modelamiento de celdas solares de perovskita mediante SCAPS-1D

El objetivo principal de este estudio es modelar y optimizar celdas solares de perovskita mediante un enfoque computacional sistemático. Para ello, se empleó el software SCAPS-1D. Esta herramienta computacional es comúnmente empleada para modelar el comportamiento eléctrico y óptico de las celdas solares, específicamente, para evaluar métricas fundamentales tales como las curvas J-V, las eficiencias energéticas, voltaje de circuito abierto V_{oc} , corrientes de circuito cerrado y los factores de llenado.

La elección de SCAPS-1D se fundamenta en su capacidad para simular dispositivos multicapa, la integración de parámetros físicos dependientes de la temperatura y condiciones de iluminación y su validación experimental reportada en varios estudios [65]. Teniendo en cuenta lo mencionado anteriormente. En esta sección se describen la arquitectura de la celda solar modelada y los parámetros de los materiales empleados en las simulaciones.

3.1.1 Diseño de celdas solar de perovskita de múltiples capas

Las celdas solares de perovskita estudiadas en este trabajo están compuestas por cinco capas tal como se ilustra en la Figura 3.1:

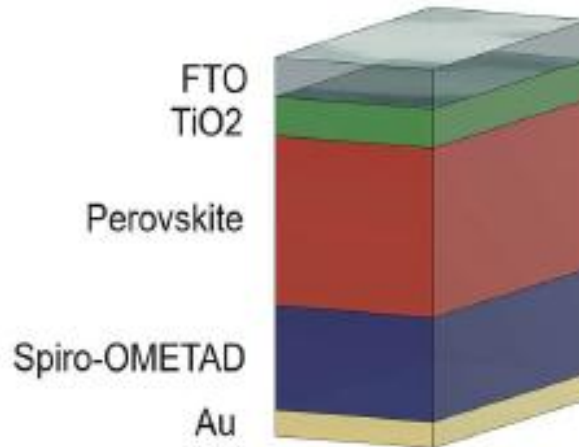


Figura 3-1 Arquitectura típica de celda solar de perovskita de cinco capas. En el esquema se ilustra la distribución de las diferentes capas que componen una celda de perovskita tradicional [42].

- La primera capa del diseño propuesto consiste en el contacto FTO. Esta capa está compuesta por óxido de estaño dopado con flúor ($\text{SnO}_2:\text{F}$). Esta capa cumple una doble función crítica, ya que en primer lugar actúa como colector de electrones. Esto se debe a

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

su alta conductividad eléctrica ($\leq 7 \times 10^{-4} \Omega \cdot \text{cm}$), lo cual permite la extracción eficiente de portadores hacia el circuito externo. En segundo lugar, este material presenta una transmitancia $> 85\%$ en el espectro visible (400-800 nm), minimizando pérdidas por reflexión [66].

- La segunda es la capa de transporte de electrones (ETL), posicionada entre el FTO y la perovskita (Figura 3-1, capa verde), utiliza dióxido de titanio (TiO_2 , fase anatasa) u óxido de estaño (SnO_2) como materiales óptimos. Ambos son semiconductores tipo *n* con una banda prohibida de 3.2 eV (TiO_2) y 3.6 eV (SnO_2), respectivamente [67].
- La tercera es la capa activa de perovskita (PVK) (Figura 3-1, capa roja). En este trabajo se considerarán FAPbI_3 para celdas con plomo y MASnI_3 como alternativa eco-amigable. Esta es la capa absorbente de perovskita, y es crucial para generar pares electrón-hueco mediante la absorción de la luz solar.
- La cuarta es la capa de transporte de huecos (HTL) (Figura 3-1, capa azul). En este caso se empleó el material Spiro-OMeTAD (una banda prohibida 3.0 eV), el cual fue seleccionado por su alineación energética con la perovskita[68].
- El contacto Au es la capa final de la celda, está hecha de oro y actúa como electrodo metálico.

Cada material fue definido en SCAPS-1D especificando los parámetros clave tales como espesor (en nanómetros), banda prohibida (E_g), afinidad electrónica (χ), permitividad relativa (ϵ_r), densidades de estados (N_c y N_v), concentraciones de dopaje (N_D o N_A), movilidad de portadores (μ_n , μ_p), y densidades de defectos (N_t) tanto en el volumen como en las interfaces. Además, se incluyeron modelos de captura de portadores (sección eficaz), tipo de defectos (neutrales o cargados) y su distribución energética (gaussiana o discreta).

Para asegurar la validez de las simulaciones, se adoptaron parámetros extraídos de la literatura científica especializada y calibraciones previas reportadas en trabajos experimentales y computacionales [69][70]. La simulación de cada celda permitió obtener la curva J–V (corriente-densidad vs. voltaje), a partir de la cual se calcularon los indicadores de desempeño eléctrico tales como J_{sc} , V_{oc} , FF y PCE. Este enfoque permitió evaluar el impacto del tipo de material absorbente (con o sin plomo) y establecer comparaciones cuantitativas entre diferentes configuraciones, sirviendo como base para el posterior análisis estadístico y modelado predictivo.

3.1.2 Simulación de propiedades ópticas y recombinación

Uno de los aspectos clave en la operación de una celda solar es su capacidad para absorber luz y convertirla en electricidad. En este sentido, SCAPS-1D permite simular diversos parámetros ópticos que influyen directamente en el desempeño del dispositivo, tales como el coeficiente de absorción, la reflexión de la luz, la generación de pares electrón-hueco y los mecanismos de recombinación.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

El modelo óptico que implementa SCAPS-1D parte de la definición del coeficiente de absorción espectral $\alpha(\lambda)$, el cual depende de la energía del fotón incidente ($h\nu$) y del ancho de banda prohibida del material. De forma práctica, el software permite ingresar directamente una tabla de absorción o emplear esta aproximación cuando no se dispone de datos experimentales detallados. Esta función de absorción define qué fracción de la radiación incidente es capaz de generar portadores de carga en cada punto del dispositivo. Además de la absorción, el software simula condiciones de iluminación que afectan la generación de portadores. En este trabajo, se utilizó como condición por defecto el espectro AM1.5G, que corresponde a la radiación solar promedio que llega a la superficie terrestre cuando el sol está a $48,2^\circ$ sobre el horizonte. Esta es una condición estándar en el análisis de celdas solares, ya que representa razonablemente bien la exposición solar en latitudes medias [71]. Dado que el efecto fotovoltaico es el fenómeno central de este estudio, el software posibilita realizar simulaciones bajo diferentes condiciones de iluminación, abarcando tanto escenarios de oscuridad como de alta luminosidad (véase Figura 3-2).

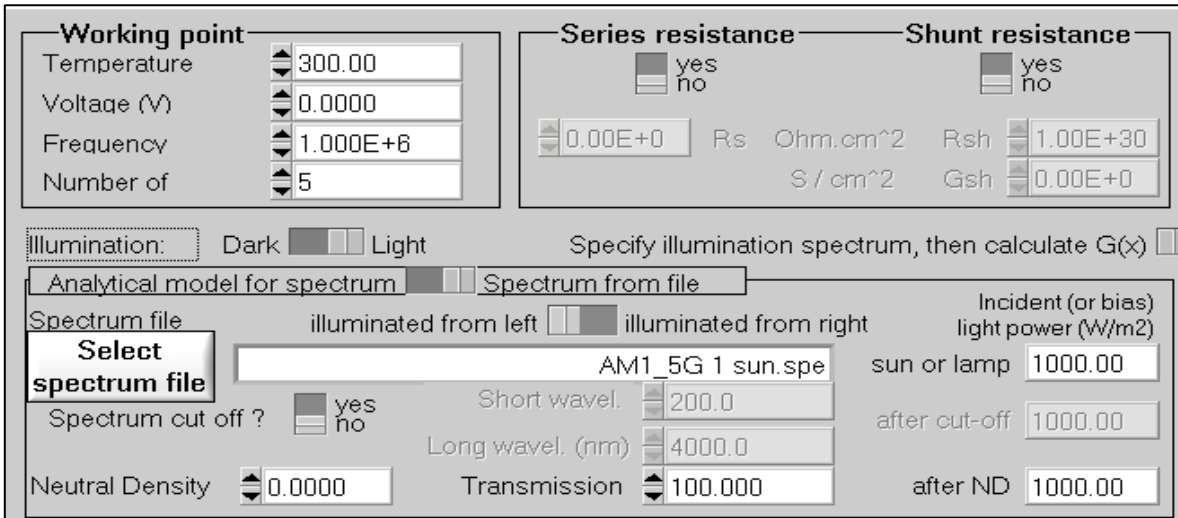


Figura 3-2. Esquema de panel de control de SCAPS en donde se configuran el punto de trabajo, tipo de iluminación y la configuración de la resistencia.

En SCAPS-1D se utilizó el panel de control para establecer las condiciones base de simulación, como se muestra en la Figura 3-2. Se fijaron cuatro parámetros clave:

- **Temperatura:** Se fijó en 300 K (27°C), condición estándar de laboratorio, ya que temperaturas superiores pueden incrementar los mecanismos de recombinación y reducir la eficiencia del dispositivo.
- **Voltaje:** Se estableció en 0 V, lo que permite evaluar el comportamiento de la celda en cortocircuito, una condición fundamental para calcular la corriente de cortocircuito (J_{sc}).

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

- **Frecuencia:** Se mantuvo el valor por defecto de 1 MHz, ya que en este estudio no se realizaron análisis capacitivos (como C-V o C-F), por lo que no se modificó esta configuración.
- **Número de puntos (pasos):** Se estableció en 5, lo cual, aunque representa una resolución baja para curvas J-V, permitió reducir la carga computacional. Esta decisión fue especialmente relevante dado que se ejecutaron 1000 simulaciones por celda, lo que requería optimizar los tiempos de cómputo sin comprometer significativamente la precisión de los resultados [72].

Un aspecto muy importante a considerar es el modelo de recombinación Shockley-Read-Hall (SRH) usado en este trabajo que explica cómo los electrones y huecos se pierden al encontrarse en defectos o impurezas dentro del material. Estos defectos actúan como "puentes" que facilitan que un electrón y un hueco se unan y desaparezcan, perdiéndose así la energía que podrían haber aportado. SCAPS usa este modelo por defecto porque es el que mejor representa lo que pasa en materiales reales, donde siempre hay imperfecciones. Además, es muy flexible: permite ajustar fácilmente la cantidad y tipo de defectos para que la simulación se parezca más a un dispositivo real. Por eso, es el punto de partida ideal para estudiar el rendimiento de una celda solar.

3.2 Metodología de simulación con el uso de SCAPS-1D

Una vez definida la arquitectura base del dispositivo fotovoltaico, se procedió a implementar la metodología de simulación en el entorno SCAPS-1D. Para este estudio, se seleccionaron dos tipos de celdas solares de perovskita que difieren en el material utilizado como capa absorbente: FAPbI₃, que contiene plomo, y MASnI₃, una alternativa libre de plomo. Ambas configuraciones comparten las mismas capas de transporte de carga: TiO₂ como capa de transporte de electrones (ETL) y Spiro-OMeTAD como capa de transporte de huecos (HTL).

Tabla 3-1. Parámetros empleados para simular la celda solar sin plomo MASnI₃ [69].

Propiedad de material	TiO ₂	MASnI ₃	Spiro OMeTAD
Espesor (nm)	50	1000	100
Banda prohibida (eV)	3.20	1.23	2.9
Afinidad electrónica (eV)	4.10	4.2	2.2
Permitividad relativa	9.0	8.2	3.0
Banda de conducción efectiva (1/cm ³)	2.2×10^{18}	1.0×10^{18}	2.5×10^{18}
Banda de valencia efectiva (1/cm ³)	1.0×10^{19}	2.9×10^{18}	1.8×10^{19}
Movilidad de electrones (cm ² /Vs)	20	1.60	2×10^{-4}
Movilidad de huecos (cm ² /Vs)	20	1.60	2×10^{-4}
Densidad de donores (1/cm ³)	1×10^{18}	0	0
Densidad de aceptores (1/cm ³)	0	1.3×10^{17}	1.3×10^{18}
Densidad de defectos (1/cm ³)	10^{15}	1.3×10^{14}	10^{15}

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

En la Tabla 3-1, se presentan los valores utilizados para la configuración sin plomo (MASnI₃) Por otro lado, en la Tabla 3-2, se resumen los parámetros para la celda con FAPbI₃, esta configuración fue seleccionada con base en la Ref. [70], en la cual se reportó un rendimiento de este tipo de celdas del 21%.

Tabla 3-2. Parámetros empleados para simular la celda solar con plomo FAPbI₃ [70].

Propiedad de material	TiO ₂	FAPbI ₃	Spiro OMeTAD
Espesor (nm)	70	350	200
Banda prohibida (eV)	3.5	1.51	2.9
Afinidad electrónica (eV)	4.0	4.0	2.2
Permitividad relativa	9.0	6.6	3.0
Banda de conducción efectiva (1/cm ³)	2.2×10^{17}	1.2×10^{19}	2.5×10^{18}
Banda de valencia efectiva (1/cm ³)	2.2×10^{17}	2.9×10^{18}	2.2×10^{18}
Movilidad de electrones (cm ² /Vs)	20	2.70	1×10^{-4}
Movilidad de huecos (cm ² /Vs)	10	1.80	1×10^{-4}
Densidad de donores (1/cm ³)	1×10^{15}	1.3×10^{16}	0
Densidad de aceptores (1/cm ³)	0	1.3×10^{16}	1.3×10^{18}
Densidad de defectos	10^{18}	4×10^{13}	10^{15}

La Figura 3-3 presenta el modelo esquemático para las configuraciones simuladas. Estos esquemas reflejan la arquitectura multicapa típica de una celda solar de perovskita en SCAPS-1D. En ambos casos, se destacan:

- **Azul:** Capa de transporte de electrones (ETL).
- **Rojo:** Capa de transporte de huecos (HTL).
- **Verde:** Capa activa de perovskita (FAPbI₃ o MASnI₃, según sea el caso)

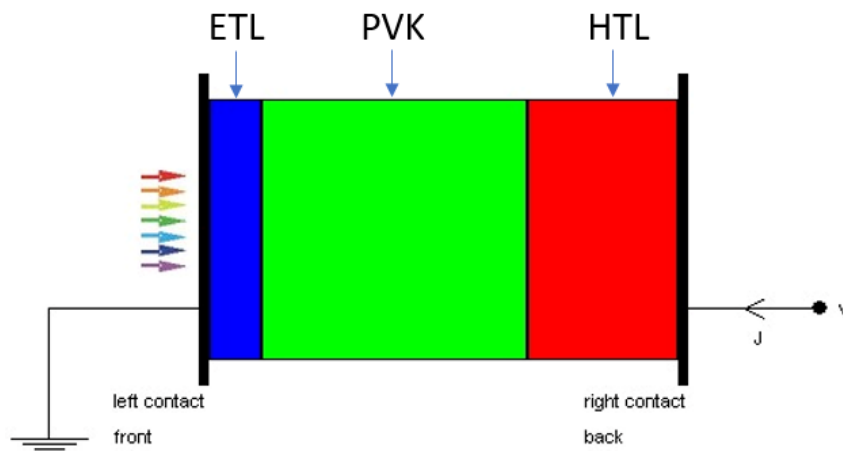


Figura 3-3. Modelo esquemático empleado para simular las celdas de perovskita.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Los contactos eléctricos están representados a la izquierda (contacto frontal, conectado a tierra y receptor de luz) y a la derecha (contacto posterior, conectado al voltaje V y encargado de la medición de corriente J). Las flechas que ingresan desde el lado izquierdo ilustran la radiación solar incidente, responsable de la generación de pares electrón-hueco en la capa activa.

3.3 Extracción y tratamiento de datos

El análisis de resultados en una simulación computacional no depende únicamente del diseño del modelo, sino también de la calidad y organización de los datos obtenidos. Por esta razón, se implementó una estrategia robusta para la extracción, filtrado y estructuración de datos, con el fin de garantizar que la información analizada reflejara de manera precisa el desempeño de las celdas solares modeladas. La Figura 3-4 presenta un diagrama de flujo explicando la metodología del proceso de extracción automática de datos.

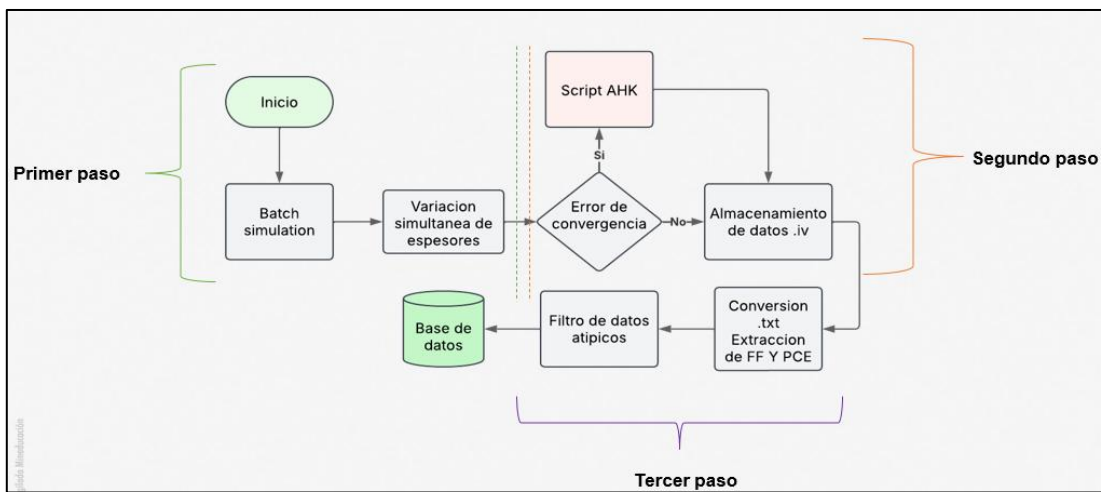


Figura 3-4 diagrama de bloques proceso de extracción automática de datos

Teniendo en cuenta el esquema mostrado en la figura 3-4, nuestra metodología para la extracción automatizada de datos se realiza simulando nuestra celda solar en la opción batch que tiene el simulador SCAPS-1D mencionado previamente en la sección 3.2.

Con el fin de analizar múltiples combinaciones de diseño, se ejecutaron 1000 simulaciones por configuración de celda (una con FAPbI₃ y otra con MASnI₃). Dado este volumen de datos, fue necesario automatizar completamente el proceso de extracción mediante tres pasos, con el fin de asegurar eficiencia, precisión y trazabilidad.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Generación de simulaciones en modo Batch

SCAPS-1D incluye un módulo llamado Batch Setup, que permite automatizar la ejecución de simulaciones múltiples con variaciones definidas de antemano. Este módulo permite explorar de forma sistemática el espacio de diseño del dispositivo, variando simultáneamente los espesores de la capa de transporte de electrones (T_{ETL}), la capa absorbente (T_{PVK}) y la capa de transporte de huecos (T_{HTL}). Cada espesor fue variado en un rango de $0.050 \mu\text{m}$ a $0.900 \mu\text{m}$ con incrementos definidos en 10 pasos. Esta configuración dio lugar a 1000 combinaciones únicas por tipo de celda, facilitando un análisis exhaustivo sin requerir ajustes manuales.

La Figura 3-5 muestra el panel de configuración del modo batch, donde se agregó el parámetro de interés thickness [μm] para cada una de las capas mencionadas. SCAPS-1D se encargó de ejecutar las simulaciones en lotes, evaluando automáticamente el desempeño del dispositivo para cada combinación de espesores. El simulador varió de manera sistemática los espesores de las capas funcionales, y finalmente generó una gran cantidad de archivos de salida con extensión “.iv”, donde cada uno de estos archivos contiene los resultados eléctricos calculados para una combinación específica de espesores.




Figura 3-5. Panel de configuración del modo Batch set-up.

Manejo de errores y validación con AutoHotKey (AHK)

La ejecución de simulaciones por lotes, SCAPS-1D puede generar errores de convergencia cuando una configuración no logra una solución estable. En estos casos, el software muestra ventanas emergentes (por ejemplo, “While calculating I(V)”), las cuales detienen el proceso hasta recibir una acción del usuario.

Para evitar interrupciones, se implementó un script en AutoHotKey (AHK), un lenguaje de automatización para Windows. Este script simula la pulsación automática de la tecla “Enter”, cerrando dichas ventanas y permitiendo que el simulador continúe con la siguiente simulación sin intervención humana. Es importante aclarar que este procedimiento no elimina los resultados erróneos, sino que permite completar el lote de simulaciones, por ello, luego se realizó una etapa de filtrado posterior para excluir de la base de datos todos los registros con valores nulos, cero o inconsistentes, asegurando así que solo se conservaran los datos técnicamente válidos y útiles para el análisis.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Extracción de resultados mediante Python

Una vez generados los archivos “.iv”, se desarrolló un script en Python, ejecutado en el entorno Spyder, que convierte cada archivo “.iv” a formato “.txt”, localiza los valores de FF y PCE, y los organiza en una base de datos estructurada que servirá como entrada para los modelos de optimización mediante aprendizaje automático.

Cada vez que se ejecuta una simulación en SCAPS-1D, el software genera una curva de densidad de corriente vs voltaje (curva J–V) como la que se ilustra en la Figura 3-6, desde la cual se derivan cuatro parámetros fundamentales: V_{oc} , J_{sc} , FF y PCE. En el contexto de este trabajo, se priorizó la recolección y análisis de los valores de FF y PCE, dado que estos permiten evaluar de forma directa el impacto de los parámetros geométricos (como los espesores de las capas) sobre el desempeño eléctrico del dispositivo.

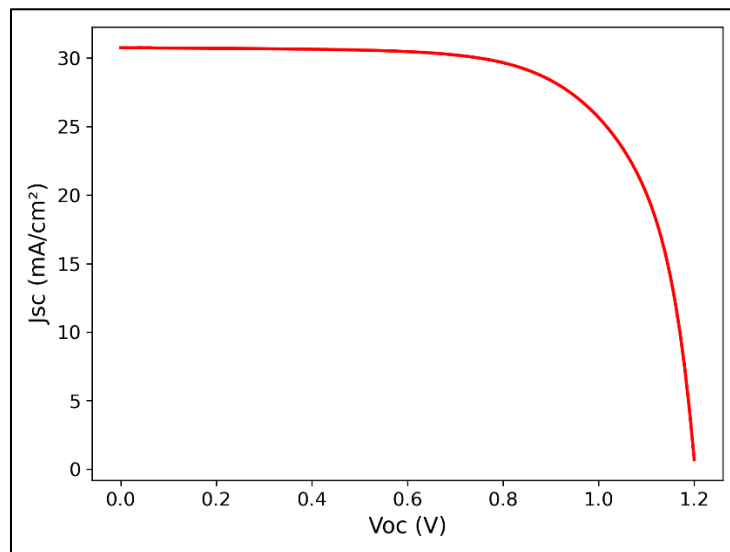



Figura 3-6. Curva de respuesta eléctrica J-V de la celda solar.

Las bases de datos utilizadas en este estudio han sido publicadas en acceso abierto mediante el repositorio OSF (Open Science Framework) con identificador digital DOI [10.17605/OSF.IO/97F84](https://doi.org/10.17605/OSF.IO/97F84) [73], con el objetivo de fortalecer la transparencia y la reproducibilidad de los resultados.

3.4 Modelo de optimización mediante aprendizaje de máquina

Se hace uso de la técnica de aprendizaje de máquinas llamada Árboles de Decisión y Random Forest debido a su capacidad de interpretar la base de datos. Esto nos permite visualizar cómo se toman las decisiones basadas en los espesores (T_{ETL} , T_{PVK} , T_{HTL}), lo que resulta útil para identificar qué características afectan más FF y la PCE. Esta técnica de optimización también nos suple la necesidad de trabajar con resultados lineales y no lineales, teniendo en cuenta

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

que las relaciones entre los espesores y el factor de llenado o la eficiencia energética no es lineal. Este modelo es capaz de entrenar múltiples árboles de decisión con diferentes subconjuntos de datos y promediar los resultados, ayudándonos a reducir el sobreajuste que un solo árbol de decisión pueda generar y aumentando la precisión del modelo. La Figura 3-7 muestra el diagrama ejemplo del árbol de decisión base para nuestro modelo de optimización:

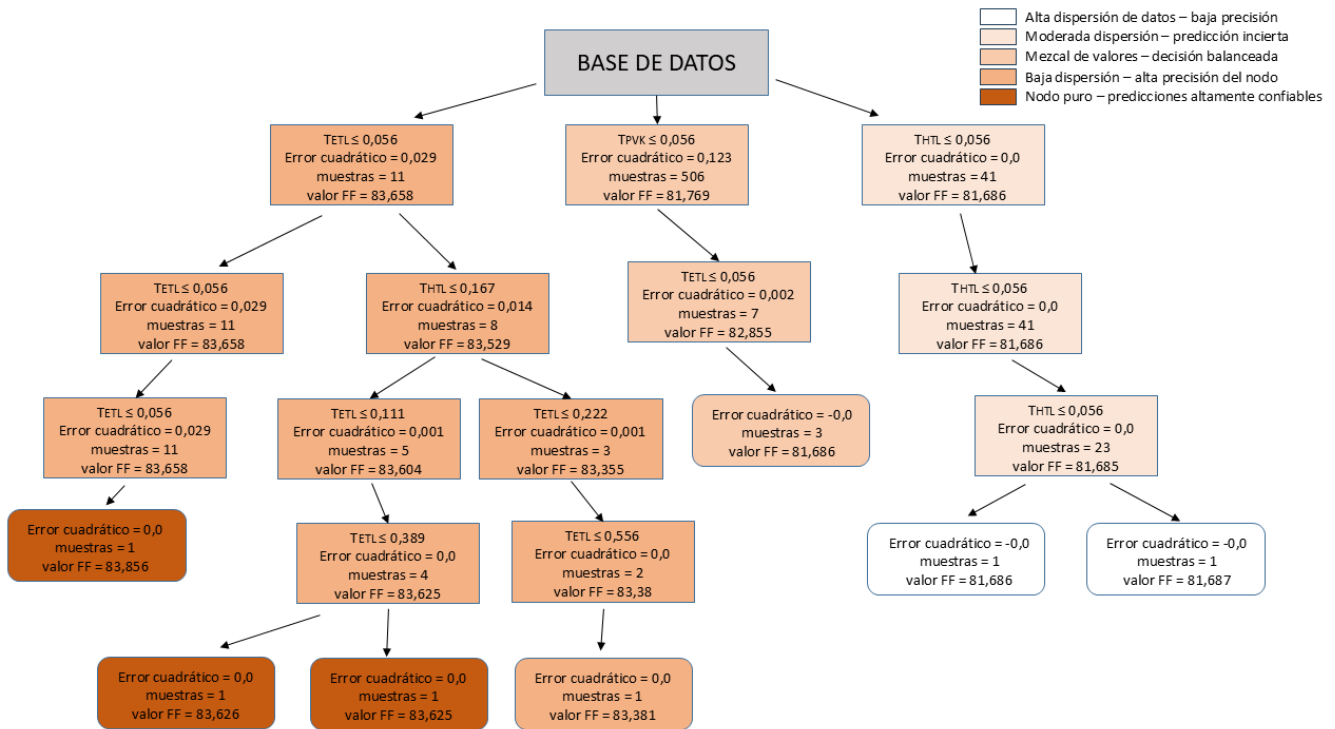



Figura 3-7. Esquema del modelo basado en árbol de decisión empleado.

En la Figura 3-7 cada cuadro del árbol de decisión representa un nodo donde el modelo evalúa una condición sobre una de las variables de entrada, como por ejemplo la condición “ $T_{PVK} \leq 0.056 \mu\text{m}$ ” indica cómo se dividen los datos: si se cumple, el árbol sigue la rama izquierda; si no, la derecha. En el mismo cuadro, la línea “error cuadrático” muestra el error cuadrático medio, el cual refleja qué tan homogéneos son los valores en ese nodo (un error más bajo indica mejor precisión). La línea “muestras” indica cuántas muestras del conjunto de datos llegan a ese punto del árbol, lo que permite entender el peso estadístico de esa decisión. Finalmente, la línea “valor FF” representa el valor promedio que predice el modelo en ese nodo, por ejemplo, un valor aproximado de FF. En las hojas del árbol (nodos finales), este valor es la predicción que ese árbol aporta al conjunto del Random Forest.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Entonces la forma en la que se adapta esto a Random forest es creando múltiples árboles de decisión con diferentes combinaciones de datos, además cada árbol prioriza distintos espesores y toma caminos diferentes, al final, el modelo realiza un promedio o vota los resultados de todos los árboles para obtener la mejor predicción de FF y PCE.

3.4.1 Algoritmo Random Forest para optimización de celda solar

Para la explicación del algoritmo trabajaremos el modelo bloque por bloque como se ilustra en la Figura 3-8.

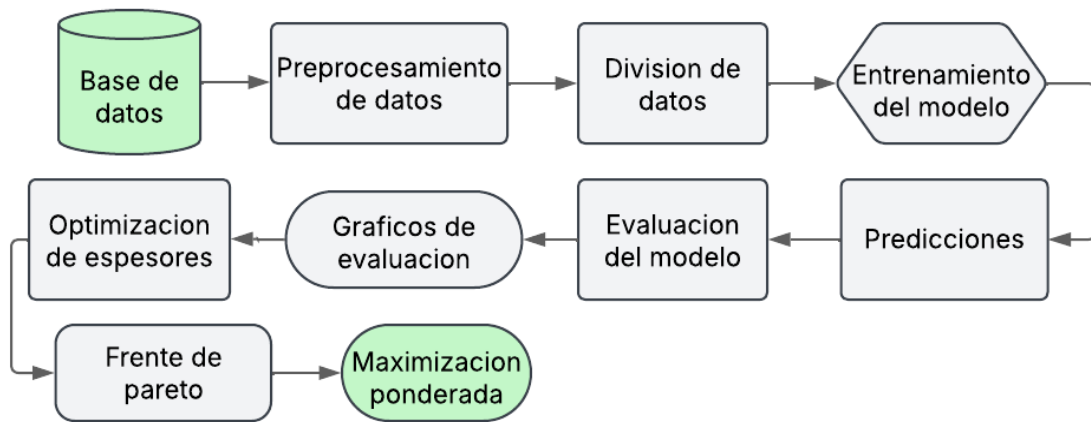


Figura 3-8 Diagrama de flujo del algoritmo aplicado.

Nuestro algoritmo de optimización parte de la base de datos creada anteriormente, donde buscamos cargar nuestro archivo desde el sistema local al entorno de ejecución, permitiéndonos leer la base de datos y luego cargarla en un dataframe de PANDAS. Nuestro pre-procesamiento de datos extrae las columnas de espesores (T_{ETL} , T_{PVK} y T_{HTL}), que son las variables independientes que se usaran como entrada del modelo, y también extrae las variables que buscamos predecir (FF y PCE), estas son dependientes de los espesores de la celda solar.

Posteriormente, la división de los datos se encarga específicamente de dividir los datos en conjuntos de entrenamiento y conjuntos de prueba. Para realizar esta función usamos un 80% como datos de entrenamiento y un 20% como datos de prueba, lo que nos permite entrenar el modelo con una parte de los datos y así evaluar su desempeño con otros datos que no se han visto antes. Para entrenar el modelo se crean dos modelos (Random Forest Regressor), uno para el FF y otro para la PCE. Este algoritmo se basa en la construcción múltiple de árboles de decisión, donde cada árbol se entrena o “aprende” con una muestra aleatoria del conjunto de entrenamiento, más conocido como técnica bootstrap.

Buscamos evaluar el modelo para medir el rendimiento de este, usando las métricas de desempeño mencionadas en el capítulo 2, tales como R^2 , MSE y RMSE. Aunque los gráficos

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

de evaluación son productos derivados del modelo, se emplean como herramienta de retroalimentación para analizar el proceso y los resultados obtenidos. Esto nos permite identificar cuáles de los tres espesores considerados ejercen mayor influencia sobre los parámetros de interés, FF y PCE. Asimismo, la comparación entre los valores estimados por el modelo y los valores simulados permite evaluar su capacidad de generalización y su coherencia con el comportamiento real del sistema, lo cual es fundamental para determinar posibles mejoras en la precisión del modelo.

Optimización multiobjetivo mediante frente de Pareto

La optimización de espesores busca identificar las combinaciones de las capas ETL, perovskita y HTL que elevan a la vez el FF y la PCE. Para ello, se construyó una malla de valores dentro de rangos factibles y, en cada punto, el modelo Random Forest entrenado predijo FF y PCE. A continuación, se aplicó el criterio de Pareto para identificar las configuraciones no dominadas, es decir, aquellas en las que ninguna de las dos métricas puede incrementarse sin disminuir la otra.

En la Figura 3-9 se presenta el diagrama de dispersión del FF en función de la PCE en el que las seis soluciones no dominadas se destacan como puntos rojos. El frente de Pareto obtenido en este estudio evidencia un comportamiento decreciente, es decir, al moverse hacia mayores valores de PCE, el FF tiende a disminuir, lo cual refleja el compromiso inherente entre ambas métricas. Como ejemplo, la Tabla 3-3 indica las coordenadas de las soluciones pertenecientes al frente de Pareto mostrado en la Figura 3-9 (puntos rojos en la figura).

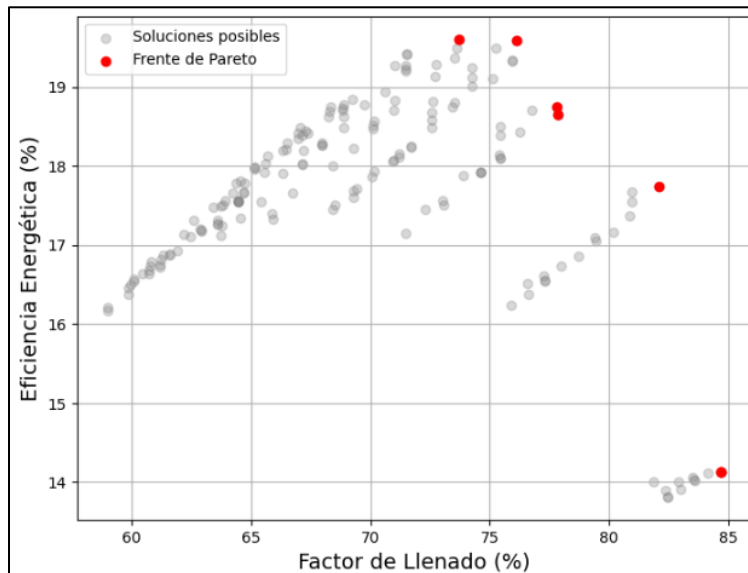


Figura 3-9. Relación FF en función de la PCE: frente de Pareto de soluciones optimas

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Tabla 3-3 Coordenadas de las soluciones no dominadas en el frente de Pareto (FF vs PCE)

FF (%)	PCE (%)	T_{ETL} (μm)	T_{PVK} (μm)	T_{HTL} (μm)
73.70	19.604	0.082692	0.5222	0.050
76.112	19.587	0.213462	0.551262	0.050
77.815	18.749	0.0500	0.5222	0.050
77.832	18.644	0.115385	0.754692	0.050
82.111	17.746	0.082692	0.435015	0.050
84.684	14.132	0.180769	0.173462	0.050

Maximización ponderada de FF y PCE

Para complementar el análisis del frente de Pareto, se formuló una función objetivo escalar que integra las dos métricas de desempeño de la celda solar mediante un parámetro de ponderación α :


$$F(T_{ETL}, T_{PVK}, T_{HTL}; \alpha) = \alpha FF(t) + (1 - \alpha)PCE(t), \alpha \in [0,1] \quad (3-1)$$

Aquí, $t=(T_{ETL}, T_{PVK}, T_{HTL})$ denota el vector de espesores de las capas, y α regula la prioridad relativa: valores de α próximos a 1 priorizan FF, mientras que valores cercanos a 0 priorizan PCE. Para poder comprender el impacto de α en el proceso de optimización de las celdas estudiadas, se evaluó la maximización de FF para seis valores representativos de α distribuidos uniformemente en el intervalo $[0, 1]$. Para cada α , se identificó la combinación de espesores que maximiza FF sobre la misma malla empleada en el análisis de Pareto. Los espesores óptimos y los resultados de FF y PCE correspondientes en cada caso son resumidos en la Tabla 3-4.

Tabla 3-4 Optimización ponderada de PCE y FF en función del parámetro α .

α	T_{ETL} (μm)	T_{PVK} (μm)	T_{HTL} (μm)	FF (%)	PCE (%)
0	0.115	0.493	0.050	73.701	19.604
0.1	0.050	0.289	0.050	79.258	19.275
0.3	0.050	0.289	0.050	79.258	19.275
0.5	0.050	0.202	0.050	82.111	17.803
0.8	0.050	0.144	0.050	85.138	14.513
1	0.050	0.144	0.050	85.138	14.513

Los resultados de la Tabla 3-4 revelan un claro balance entre eficiencia y factor de llenado al variar el parámetro de ponderación α . Para $\alpha=0$, la función objetivo prioriza exclusivamente la PCE, lo que conduce a un espesor relativamente grueso de perovskita (0.493 μm) y un FF moderado igual a 73.7%. A medida que α aumenta, el algoritmo favorece configuraciones con capas más delgadas, en especial en la perovskita, elevando progresivamente el FF mientras sacrifica rendimiento energético. Por ejemplo, cuando α toma un valor igual a 0.8

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

se alcanza un máximo de 85.14% para el FF, sacrificando la PCE, esta última se reduce hasta un 14.5 %. Por otro lado, el caso intermedio ($\alpha=0.5$) demuestra un equilibrio óptimo, con un FF de 82.11 % y una PCE de 17.80 %, lo que sugiere que una ponderación equitativa produce un diseño de espesores que equilibra adecuadamente ambas métricas. Este análisis cuantitativo confirma que, mediante la selección adecuada de α , es posible dirigir el proceso de optimización hacia soluciones con perfiles de rendimiento ajustados a requisitos específicos de estabilidad, costo o aplicación final.

Herramienta interactiva de estimación

La “Calculadora” interactiva se desarrolló como un módulo Python autónomo que complementa el proceso de optimización al proporcionar estimaciones instantáneas de desempeño para cualquier combinación de espesores de capas en celdas solares de perovskita. Su interfaz solicita primeramente la elección del material activo (FAPbI₃ o MASnI₃), con lo cual se cargan automáticamente los modelos Random Forest entrenados para cada composición. A continuación, se ingresan los espesores de ETL, perovskita y HTL, que se normalizan automáticamente con las mismas transformaciones estadísticas utilizadas previamente.

Con los datos normalizados, la Calculadora invoca los modelos entrenados (RF_FF.pkl y RF_PCE.pkl) para predecir el factor de llenado y la eficiencia de conversión de potencia. La Figura 3-10 presenta el diagrama de flujo, donde se muestran los pasos: selección de material, normalización de entradas, predicción de parámetros eléctricos y despliegue de resultados. Este enfoque elimina la necesidad de ejecutar nuevas simulaciones y facilita la validación rápida de configuraciones, la comparación de diseños y la exploración de ajustes finos en etapas avanzadas del desarrollo. Esta herramienta elimina la necesidad de ejecutar nuevas simulaciones en SCAPS u otro simulador de celdas solares de perovskita y facilita la validación rápida de configuraciones, la comparación de diseños y la exploración de ajustes finos en etapas avanzadas del desarrollo.

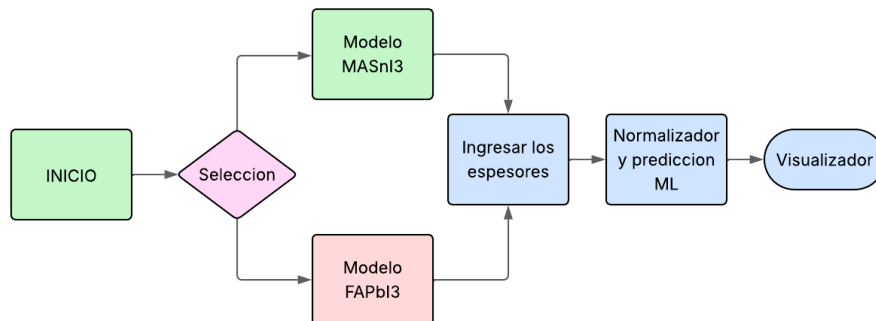



Figura 3-10. Flujo funcional del algoritmo estimador predictivo de FF y PCE para celdas solares de perovskita

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Optimización de hiperparámetros del modelo de ML

Como parte del ajuste de hiperparámetros, se realizó una búsqueda aleatoria (RandomizedSearch) para optimizar el modelo de predicción PCE. Esta búsqueda considero la distribución definida de hiperparámetros (param_grid_rf) y se realizaron 400 iteraciones aleatorias, utilizando validación cruzada de 5 particiones (cv=5). La selección del mejor modelo se basó en la métrica de error cuadrático medio (RMSE). A diferencia del enfoque con valores fijos, este procedimiento también fue aplicado al modelo correspondiente a FF, permitiendo encontrar combinaciones de hiperparámetros más adecuadas. Esto nos ayudó a obtener un balance entre precisión y eficiencia computacional, ayudándonos a disminuir el riesgo de un sobreajuste.

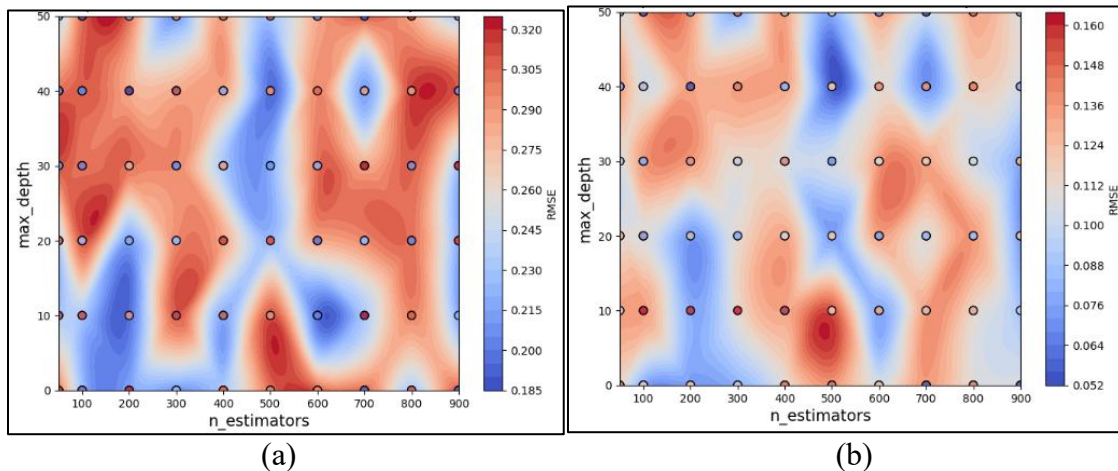


Figura 3-11. (a) Mapa de contornos con concentración de RMSE bajo para: (a) FF y (b) PCE.

El mapa de contorno de la figura 3-11(a) y figura 3-11(b) representan cómo varía el RMSE del modelo en función de dos hiperparámetros del algoritmo Random Forest: el número de árboles (n_estimator) y la profundidad máxima de cada árbol (max_depth). A medida que se exploran diferentes combinaciones de estos parámetros, se observa que los colores azules, especialmente el azul intenso, indican zonas donde el RMSE es más bajo, es decir, donde el modelo predice con mayor precisión, mientras que las zonas que reflejan el color rojo indican un mayor error, lo cual implica un peor desempeño del modelo. La figura 3-11(a) evidencia que el modelo alcanza su mejor desempeño en la predicción del FF cuando se utilizan 200 árboles y una profundidad máxima de 10, configuración en la que el RMSE se reduce a hasta 0.185 %, indicando un ajuste excelente. Por otro lado, en la figura 3-11(b) observamos que el modelo obtiene menor RMSE para el PCE cuando el modelo alcanza su punto óptimo cuando se utilizan 500 árboles y una profundidad máxima de 40, donde el RMSE es mínimo, cercano a 0.052 %.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

4. RESULTADOS Y DISCUSIÓN

En esta sección, se presentan y analizan los resultados obtenidos a partir de la simulación de celdas solares de perovskita utilizando el software SCAPS-1D, así como la validación de las predicciones realizadas mediante el modelo de aprendizaje automático Random Forest. El objetivo es evaluar la capacidad predictiva del modelo, comparando sus estimaciones con los resultados obtenidos de las simulaciones y otros datos reportados en publicaciones científicas. A través de este análisis, se busca identificar las tendencias clave en la optimización de celdas solares y analizar la efectividad del modelo para la estimación de parámetros críticos, como el factor de llenado (FF) y la eficiencia de conversión de potencia (PCE).

4.1 Resultados mediante técnicas multiobjetivo.

Con base en los datos simulados y procesados mediante el modelo de Random Forest descrito en el capítulo 3, se llevó a cabo un análisis de optimización multiobjetivo orientado a identificar configuraciones estructurales que maximicen simultáneamente el factor de llenado y la eficiencia de conversión de potencia. Para ello, se evaluó una malla tridimensional de combinaciones de espesores en las capas ETL, PVK y HTL, y se aplicó el criterio de no dominancia para construir el frente de Pareto asociado a estas dos métricas clave.

Este frente constituye el conjunto de soluciones óptimas en el sentido de Pareto, es decir, configuraciones en las que no es posible mejorar una métrica sin comprometer la otra. Como se mostró en la Tabla 3-3 y en la Figura 3-9, se identificaron seis configuraciones no dominadas que representan diferentes compromisos entre FF y PCE. Los resultados evidencian una relación inversa entre estas dos métricas en el espacio de soluciones, reflejando un comportamiento de compensación típico en problemas de diseño de dispositivos fotovoltaicos.

El análisis realizado valida la formulación del problema como un caso de optimización multiobjetivo con funciones objetivo acopladas y no lineales, en el que las variables de decisión (espesores de ETL, PVK y HTL) fueron restringidas a rangos físicamente viables definidos en la sección 3.2. La aplicación del modelo predictivo entrenado en el capítulo 3.4 permitió evaluar rápidamente más de mil configuraciones, evitando el uso reiterado de simulaciones en SCAPS-1D, lo cual fue crucial para alcanzar una cobertura densa del espacio de diseño.

Además del análisis puramente basado en Pareto, se implementó una estrategia de maximización ponderada (ver ecuación 3-1), en la cual se introdujo un parámetro α que permite ajustar la prioridad relativa entre FF y PCE. Este enfoque permite recorrer el frente

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

de Pareto modificando los pesos asignados a cada función objetivo. La Tabla 3-4 resume las configuraciones óptimas obtenidas para diferentes valores de α , ilustrando cómo el ajuste de prioridades permite explorar distintas soluciones, desde aquellas que maximizan la PCE hasta las que privilegian un mejor comportamiento FF.

La Figura 4-1 muestra los resultados obtenidos mediante la aplicación del algoritmo con maximización ponderada para una configuración con espesores iguales a 0,05 μm , 0,2025 μm y 0,05 μm en las capas de TTL, perovskita y HTL, respectivamente. Con ello, se evidencia la capacidad de la técnica propuesta para identificar configuraciones de espesores que optimizan simultáneamente PCE y FF.

```

Valores óptimos para maximización ponderada (alpha=0.5):
Espesores: T_ETL    0.050000
T_PVK    0.202523
T_HTL    0.050000
Name: 2, dtype: float64
FF: 82.111
PCE: 17.803

```


Figura 4-1. Formato de salida, visualización de resultados para optimización de celda solar MASnI_3 .

4.2 Análisis de los resultados para celdas solares de FAPbI_3

El primer paso en la validación de la capacidad del modelo Random Forest para predecir correctamente el factor de llenado y la eficiencia de conversión de potencia a partir de los espesores de las tres capas clave de la celda solar (HTL, PVK y ETL), consistió en comparar los resultados obtenidos con los que se generan utilizando el software SCAPS-1D. La comparación se realizó seleccionando una configuración estándar de celda solar de perovskita, específicamente una celda basada en FAPbI_3 , cuyas dimensiones fueron tomadas de la referencia [70].

- **Espesor de la capa ETL:** 0,050 μm
- **Espesor de la capa PVK:** 0,600 μm
- **Espesor de la capa HTL:** 0,200 μm

La simulación de esta configuración en SCAPS-1D arrojó un FF de 83.47% y una PCE de 26.04%, tal como se observa en la parte inferior de la simulación de la curva J-V en SCAPS-1D, como se muestra en la Figura 4-2(a). Estos espesores fueron luego ingresados en el modelo predictivo de Random Forest propuesto, utilizando los mismos espesores como parámetros de entrada. Con ello, el modelo logró las estimaciones de FF de 83,66 % y una PCE de 25.65%, como se muestra en la Figura 4-2(b).

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

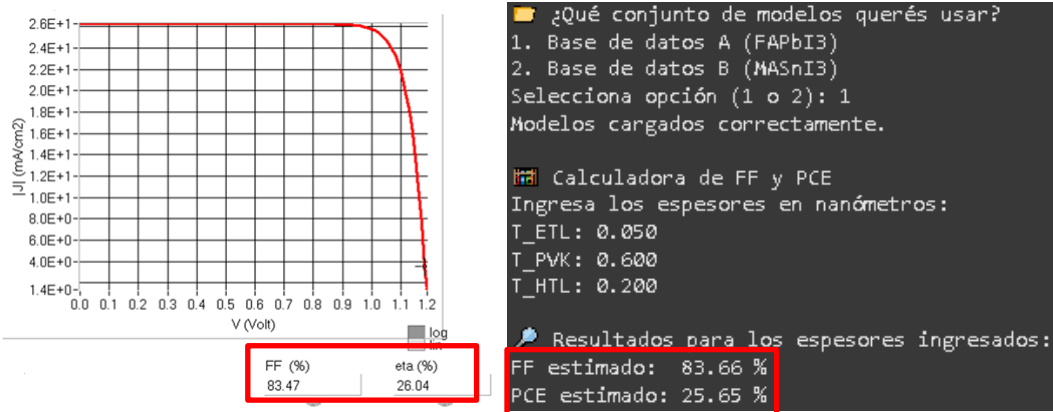


Figura 4-2. (a) Resultado curva J-V de SCAPS-1D. (b) Formato de salida, visualización de resultados modelo ML.

Los errores relativos entre los resultados obtenidos por SCAPS-1D y el modelo de Random Forest propuesto son:

- **Error porcentual en FF:**

$$Error\ FF\ (\%) = \frac{|FF_{SCAPS} - FF_{ML}|}{FF_{SCAPS}} \times 100\% = 0.22\%$$

- **Error porcentual en PCE:**

$$Error\ PCE\ (\%) = \frac{|PCE_{SCAPS} - PCE_{ML}|}{PCE_{SCAPS}} \times 100\% = 1.49\%$$

Estos resultados muestran una alta concordancia entre las predicciones del modelo y los resultados de simulación. Este comportamiento valida la capacidad del algoritmo para replicar con precisión el comportamiento eléctrico de la celda solar, destacando su utilidad para la estimación de parámetros clave sin necesidad de realizar simulaciones computacionalmente costosas.

Una vez confirmada la alta concordancia entre los resultados obtenidos mediante el modelo de Random Forest y las simulaciones realizadas en SCAPS-1D para un caso puntual, se procedió a realizar una serie de pruebas adicionales. Estas pruebas consistieron en variar los espesores de las capas ETL, PVK y HTL dentro de los rangos previamente definidos, con el objetivo de evaluar el desempeño y la robustez del modelo en un conjunto más amplio de configuraciones.

El propósito de esta fase fue evaluar el grado de correlación y estabilidad del modelo predictivo cuando se aplicaba a múltiples combinaciones de espesores, permitiendo observar cómo se comportaba en una variedad de escenarios. Este análisis permitió verificar la capacidad del modelo para generalizar a nuevos datos, y para identificar posibles patrones de desempeño a medida que los espesores de las capas variaban. A continuación, se presenta la

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO		Código	FDE 089
			Versión	04
			Fecha	24-02-2020

Tabla 4-1, en la cual se comparan los resultados obtenidos tanto por simulación como por predicción del modelo Random Forest, permitiendo comparar las estimaciones de FF y PCE para distintas combinaciones de espesores. Este enfoque proporciona una visión clara del comportamiento general del modelo, evidenciando su precisión y fiabilidad a lo largo de una gama de configuraciones.

Tabla 4-1. Errores porcentuales para FAPbI₃ (Modelo vs SCAPS- 1D)

ESPEORES (µm)			Simulado en SCAPS-1D		Calculado ML		% Error	
ETL	PVK	HTL	FF (%)	PCE (%)	FF (%)	PCE (%)	FF (%)	PCE (%)
0,050	0,600	0,200	83,47	26,04	83,66	25,60	0,22	1,68
0,100	0,400	0,350	83,04	26,48	83,65	25,65	0,73	3,13
0,400	0,700	0,500	83,25	26,05	83,66	25,30	0,49	2,61
0,300	0,900	0,501	82,82	26,37	83,0	26,37	0,21	2,87

Los resultados de la Tabla 4-1 permite concluir que, en todos los casos analizados, los valores de FF y PCE calculados por el modelo muestran una alta concordancia con los obtenidos en las simulaciones, con errores porcentuales que, no superan el 3.13 % en ninguno de los casos.

Es interesante notar que, incluso al variar significativamente los espesores de las capas, especialmente en el caso de la perovskita, donde los espesores se exploran entre 0,400 µm y 0,900 µm, el modelo mantiene una consistencia notable en sus predicciones. Por ejemplo, en la primera configuración, los errores para FF y PCE fueron de 0,22% y 1,68%, respectivamente. En la segunda y tercera configuraciones, aunque las diferencias aumentan ligeramente, siguen siendo pequeñas y dentro de un rango aceptable (inferior al 5%), especialmente considerando que el modelo no se basa en simulaciones físicas, sino en inferencias estadísticas.

4.3. Análisis de los resultados para celdas solares de MASnI₃

De manera análoga a las pruebas realizadas con las celdas solares FAPbI₃, se evaluó la capacidad del modelo Random Forest para predecir los parámetros de desempeño en celdas solares basadas en MASnI₃. En este caso, se utilizaron configuraciones de espesores diferentes para las capas ETL, PVK y HTL, basadas exclusivamente en datos simulados generados con SCAPS-1D. Esto garantiza que las comparaciones se realicen bajo condiciones homogéneas y consistentes, proporcionando una referencia clara para evaluar la precisión del modelo.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Los resultados obtenidos, que se detallan en la Tabla 4-2, muestran una excelente concordancia entre los valores de FF y PCE generados por SCAPS-1D y los estimados por el modelo Random Forest. En tres de los cuatro casos evaluados, las diferencias porcentuales entre ambos parámetros fueron mínimas, con algunas diferencias prácticamente nulas. Estos resultados demuestran que el modelo ha sido correctamente entrenado con los datos simulados, aprendiendo de manera precisa las relaciones entre los espesores de las capas y el rendimiento de la celda. Esta consistencia subraya la robustez del modelo y su capacidad para realizar predicciones confiables incluso en configuraciones diversas.

En cuanto a la variabilidad observada, esta es más notoria en configuraciones con espesores más gruesos, lo que podría ser resultado de las interacciones más complejas entre las capas en estas configuraciones extremas. Sin embargo, estas diferencias son pequeñas y no comprometen la validez del modelo, ya que los errores porcentuales siguen dentro de rangos aceptables. Por ejemplo, en la primera configuración con una capa ETL de 0,05 μm , la diferencia en FF fue de 0,13%, lo cual es aún un margen razonable.

Tabla 4-2. Errores porcentuales para MASnI_3 (Modelo vs SCAPS-1D).

ESPESTORES (μm)			SCAPS		Calculado ML		% Error	
ETL	PVK	HTL	FF (%)	PCE (%)	FF (%)	PCE (%)	FF (%)	PCE (%)
0,050	0,600	0,100	81,53	24,23	81,40	23,72	0,13	2,10
0,073	0,723	0,080	81,69	24,76	81,67	24,70	0	0,12
0,400	0,900	0,200	81,82	25,08	81,83	25,0	0,03	0
0,278	0,640	0,125	81,58	24,26	81,55	24,13	0	0,37

En general, el modelo mantiene una alta precisión en la predicción de FF y PCE, con errores porcentuales que no superan el 2,5% en la mayoría de los casos. En particular, la PCE se predijo con una exactitud destacable, con errores menores al 1% en tres de los cuatro escenarios. En uno de los casos, la diferencia fue nula (0%) en PCE, lo que valida aún más la capacidad del modelo para estimar con precisión el desempeño energético de las celdas solares.

En configuraciones más extremas, como en el caso de los espesores más grandes ($T_{\text{ETL}}=0,400 \mu\text{m}$, $T_{\text{PVK}}=0,900 \mu\text{m}$, $T_{\text{HTL}}=0,200 \mu\text{m}$), el modelo demostró su capacidad para manejar combinaciones complejas sin perder precisión. Estos resultados refuerzan la robustez del modelo, ya que las predicciones siguen siendo precisas incluso al explorar configuraciones de espesores que se desvían significativamente de los valores comunes.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

Cabe mencionar que algunos de los valores escogidos en las tablas 4-1 y 4-2 para los correspondientes espesores de las capas, han sido valores aleatorios entre 50 μm hasta 900 μm que no se encuentran en la base de datos, esto con el propósito de validar la capacidad del modelo propuesto para predecir valores por fuera de la base de datos y la capacidad de generalización o interpolación del mismo.

4.4. Comparación del rendimiento entre celdas solares basadas en FAPbI_3 y MASnI_3

Al comparar el rendimiento de las celdas solares basadas en FAPbI_3 (con plomo) y MASnI_3 (libres de plomo), se observa que las celdas con plomo presentan un desempeño superior en términos de PCE y FF. En los casos analizados en las bases de datos, las celdas FAPbI_3 alcanzaron un PCE simulado de 26,04% y un FF de 83,47%, mientras que las celdas MASnI_3 , aunque competitivas, mostraron valores ligeramente más bajos, con un PCE de 25,08 % y un FF de 81,82 % en configuraciones similares. A pesar de esta diferencia de rendimiento, las celdas MASnI_3 tienen la ventaja de ser una opción más sostenible, debido a la menor toxicidad del estaño frente al plomo, lo que las hace más adecuadas para aplicaciones a largo plazo desde una perspectiva medioambiental. Los errores porcentuales entre las predicciones del modelo Random Forest y los resultados simulados en SCAPS-1D para ambas celdas fueron pequeños, lo que valida la precisión del modelo en la estimación de FF y PCE. Estos resultados subrayan que, aunque las celdas con plomo siguen siendo más eficientes, las celdas libres de plomo representan una alternativa viable que requiere más investigación para mejorar su rendimiento y estabilidad a largo plazo.

4.5. Capacidad de generalización del modelo y comparación con resultados de la literatura

Uno de los aspectos más cruciales para evaluar la calidad de un modelo predictivo es su capacidad para generalizar, es decir, para realizar predicciones precisas incluso cuando se enfrenta a datos nuevos que no formaron parte de su conjunto de entrenamiento. En el contexto de este trabajo, esto implica comprobar si el modelo de aprendizaje automático es capaz de replicar valores de rendimiento como la PCE y el FF que se encuentran en estudios previos, basados en configuraciones reales de celdas solares de perovskita presentadas en la literatura científica.

Demostrar que el modelo puede replicar estos valores es una validación importante, ya que no solo indica que el modelo ha aprendido de manera efectiva con los datos de simulación con los que fue entrenado, sino que también ha logrado captar las relaciones profundas entre las características físicas de una celda y su rendimiento. Este nivel de generalización permite que el modelo sea utilizado para hacer predicciones en nuevas configuraciones sin necesidad de realizar simulaciones completas en SCAPS-1D cada vez, lo que representa una gran

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

ventaja en términos de tiempo y recursos. Para evaluar la capacidad de generalización del modelo, se compararon las predicciones realizadas por Random Forest con los resultados de simulaciones reportadas en varios artículos científicos. Esta comparación es fundamental para verificar si el modelo puede replicar configuraciones propuestas por otros investigadores y si sus estimaciones son suficientemente cercanas a los valores reales como para ser utilizadas de manera confiable en el análisis y validación de nuevas propuestas tecnológicas.

La Tabla 4-3 presenta los resultados comparativos para celdas solares de FAPbI₃, donde se contrastan los valores de FF y PCE reportados en cuatro publicaciones recientes, con los valores estimados por el modelo de Random Forest propuesto en esta tesis. Además, se incluyen los errores porcentuales correspondientes para cuantificar el grado de aproximación entre las simulaciones y las predicciones del modelo. Los resultados muestran que el modelo de Random Forest es capaz de replicar con alta precisión los valores de FF y PCE, con errores generalmente bajos en la mayoría de los casos. En tres de los cuatro estudios evaluados, los errores porcentuales en FF y PCE son muy pequeños, con un máximo de 1,98% en FF y 1,84% en PCE, lo que demuestra que el modelo puede adaptarse bien a configuraciones ya propuestas en la literatura.

Tabla 4-3. Comparación con resultados reportados en la literatura para celdas solares con FAPbI₃.

REFERENCIA	Reportado Autores		Calculado ML		% Error	
	FF (%)	PCE (%)	FF (%)	PCE (%)	FF (%)	PCE (%)
[74]	81,12	21,25	80,55	21,09	0,71	0,75
[70]	69,17	18,55	69,73	18,64	0,80	0,48
[75]	85,12	26,61	83,43	26,12	1,98	1,84
[76]	81,15	25,83	79,72	23,70	1,76	8,24

Sin embargo, también se observa una discrepancia más notable en el estudio de [76], donde el error en PCE alcanza el 8,24%, lo que es considerablemente mayor que en los otros casos. Esta diferencia puede explicarse por una serie de factores, como condiciones específicas de simulación empleadas en cada artículo, como tratamientos de superficie, defectos estructurales no considerados en el modelo o condiciones de frontera que podrían no estar reflejadas en la base de datos original utilizada para entrenar el modelo. Es importante destacar que, a pesar de estas discrepancias, el comportamiento general del modelo sigue siendo sólido y confiable, y la mayoría de los resultados muestran una capacidad significativa para adaptarse a configuraciones de celdas solares descritas en la literatura científica.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO		Código	FDE 089
			Versión	04
			Fecha	24-02-2020

4.6. Comparación del rendimiento de celdas solares MASnI_3 y evaluación del modelo predictivo

En la Tabla 4-4, se presentan los errores porcentuales de las celdas de MASnI_3 , comparando los resultados reportados en la literatura reciente con las predicciones realizadas por el modelo de ML propuesto. Se incluyen los valores de FF y PCE reportados en estudios recientes, y se calculan los errores relativos para cada caso. El modelo de Random Forest mostró un buen rendimiento en comparación con los valores reportados en diversos artículos científicos para celdas solares de MASnI_3 , con errores razonables en FF y PCE. Por ejemplo, en los trabajos de [69], [77], el error en FF fue menor al 4%, lo que demuestra que el modelo es capaz de predecir con alta precisión incluso en configuraciones que no fueron incluidas explícitamente en el conjunto de entrenamiento. No obstante, en el caso del estudio [78], el modelo presentó una gran diferencia en FF (12,02 %) y en PCE (2,81 %), lo que sugiere que algunas de las características particulares de esa configuración, como el espesor de las capas o los tratamientos de superficie, no fueron adecuadamente representadas en el conjunto de entrenamiento del modelo.

A pesar de las discrepancias más grandes en ciertos casos, el modelo sigue demostrando una gran capacidad de generalización, especialmente cuando se enfrenta a configuraciones más comunes o estándar. Los resultados de esta comparación confirman que el modelo Random Forest puede ser útil como herramienta predictiva preliminar para la evaluación de celdas solares de MASnI_3 , ofreciendo predicciones razonables en términos de FF y PCE sin necesidad de recurrir a simulaciones completas y costosas como las de SCAPS-1D. Aunque el modelo presenta variaciones en configuraciones más complejas, su capacidad para proporcionar estimaciones rápidas y confiables lo convierte en una herramienta valiosa para la optimización y validación de nuevos diseños en el campo de las celdas solares.

Tabla 4-4 Comparación con resultados reportados en la literatura para celdas solares con MASnI_3 .

DOCUMENTO	Reportado Autores		Calculado ML		Error	
	%FF	%PCE	%FF	%PCE	% FF	% PCE
[69]	82,65	27,20	82,0	28,26	0,78%	3,89%
[78]	57,46	6,40	50,55	6,22	12,02%	2.81%
[79]	84,50	24,14	81,80	24,71	3,19%	2,36%
[77]	82,99	16,19	80,46	16,71	3,04%	3,11%

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

4.7. Desempeño del modelo de predicción

Una vez entrenado el modelo de Random Forest, se procedió a evaluar su capacidad predictiva utilizando tres métricas estándar de regresión: el coeficiente de determinación (R^2) el error absoluto medio (MAE) y la raíz del error cuadrático medio (RMSE). Estas métricas permiten cuantificar la precisión de las predicciones realizadas por el modelo en relación con los valores reales generados mediante simulaciones en SCAPS-1D.

Metricas de evaluacion para celda solar FAPbI₃

Los resultados obtenidos fueron altamente satisfactorios para ambas variables objetivo. En el caso del FF, el modelo alcanzó un valor de $R^2 = 0.9454\%$ lo cual indica que más del 94.54 % de la variabilidad de los datos es explicada por el modelo. Además, los errores obtenidos fueron mínimos, con un MAE de aproximadamente 0.089% y un RMSE de 0.164%. Estos valores reflejan una excelente precisión en las predicciones, con diferencias prácticamente despreciables respecto a los valores reales.

De forma similar, para la PCE, el modelo alcanzó un $R^2 = 0.9261\%$, lo cual sigue siendo indicativo de un ajuste sobresaliente. Aunque los errores fueron ligeramente mayores que en el caso del FF, con un MAE de 0.025% y un RMSE de 0.038%, estos siguen siendo considerablemente bajos, especialmente considerando la complejidad no lineal que presenta la relación entre las características estructurales de la celda y su eficiencia.

Metricas de evaluacion para celda solar MASnI₃

Los resultados obtenidos para la celda solar basada en MASnI₃ evidencian un desempeño muy satisfactorio del modelo predictivo. Para el FF, se alcanzó un coeficiente de determinación R^2 de 0.9456%, lo que indica que el modelo logra explicar aproximadamente el 94.56 % de la variabilidad de los datos. Además, los errores fueron muy bajos, con un MAE de 0.0021% y un RMSE de 0.0048%, reflejando una excelente precisión en las predicciones.

En cuanto a la PCE, el modelo obtuvo un R^2 de 0.9632%, lo que representa un ajuste aún más sólido. Si bien los errores fueron algo mayores que en el caso del FF con un MAE de 0.046% y un RMSE de 0.057% estos valores siguen siendo bajos, especialmente considerando la complejidad no lineal que implica la predicción del desempeño fotovoltaico.

La calidad de los datos de entrada, generados mediante simulación controlada, garantiza consistencia y bajo nivel de ruido, lo que favorece un aprendizaje más robusto. Además, la selección adecuada de características y la exploración cuidadosa de los hiperparámetros del

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

modelo durante el entrenamiento contribuyeron de manera significativa a optimizar su desempeño.

Recursos computacionales

Para llevar a cabo el entrenamiento y validación del modelo, se utilizó un equipo de cómputo con las especificaciones resumidas en la Tabla 4-5.

Tabla 4-5 Detalle del equipo de cómputo empleado

Componente	Especificación
Procesador	AMD Ryzen 5 5500U @ 2.10 GHz
Memoria RAM	16 GB
Tarjeta grafica	AMD Radeon Graphics 8™ (496 MB)
Almacenamiento	SSD Kingston SNV2S1000G (932 GB)
Sistema operativo	Windows 10 Pro 64 bits

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020


5. CONCLUSIONES Y RECOMENDACIONES.

Durante la realización de este trabajo se implementó y validó exitosamente una metodología computacional basada en SCAPS-1D para modelar celdas solares de perovskita. La construcción de los modelos *n-i-p* permitió simular con precisión el comportamiento eléctrico de celdas solares de perovskita bajo distintas condiciones. A través de series de simulaciones se generaron 1000 combinaciones de espesores para cada tipo de perovskita, lo cual no solo confirmó la confiabilidad de SCAPS-1D para reproducir curvas J-V acordes con la literatura, sino que también facilitó la creación de una base de datos sobre la influencia de parámetros estructurales en rendimiento fotovoltaico de las mismas. De esta manera, se logró implementar y validar un modelo computacional que reflejara fielmente la respuesta real de las celdas de perovskita.

Por otro lado, la evaluación comparativa del desempeño de celdas solares con y sin plomo puso de manifiesto diferencias significativas en eficiencia de conversión y factor de llenado. Para las celdas de FAPbI₃, se registraron PCE máximas cercanas al 26.37 % y FF alrededor del 83 %, mientras que en las celdas de MASnI₃, a pesar de su mayor desafío en estabilidad por la tendencia oxidativa del estaño, se alcanzaron valores de PCE próximos al 25 % y FF alrededor del 81.83 %. Estos resultados confirman la viabilidad de las perovskitas libres de plomo como alternativa más sostenible, aunque con ligeras pérdidas de rendimiento frente a las basadas en plomo. Asimismo, el análisis detallado de la variación de espesores en la capa ETL, la capa activa y la capa HTL evidenció rangos óptimos donde se maximiza simultáneamente la PCE y el FF.

Adicionalmente, se implementó un modelo de aprendizaje de máquina basado en Random Forest para predecir simultáneamente PCE y FF como función de los espesores de las capas en la celda solar basada en FAPbI₃, el cual permitió reducir el tiempo de simulación y facilitar la exploración de diseños. La búsqueda en malla de hiperparámetros y la validación cruzada condujeron a un ajuste preciso, con coeficientes de determinación superiores a 0.9261 % para ambas métricas. El error relativo promedio fue menor al 2 % en la mayoría de los casos, lo que evidencia que los modelos capturan eficazmente los efectos no lineales de los parámetros estructurales. Con esta aproximación, se logró maximizar la eficiencia energética y el factor de llenado mediante aprendizaje de máquinas. Además, la construcción del frente de Pareto permitió identificar, de manera simultánea, las configuraciones óptimas que no sacrifican un indicador en favor del otro, ofreciendo una guía concreta para la selección de espesores en futuras etapas experimentales.

En comparación, la celda solar basada en MASnI₃ presentó un desempeño aún más favorable. Se alcanzaron niveles de ajuste superiores, con coeficientes de determinación mayores al 0.9456 % para ambas métricas y errores relativos promedio inferiores al 2.1 %, lo que refleja

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

una mayor precisión en la predicción del comportamiento fotovoltaico. Estos resultados consolidan la eficacia del modelo para materiales sin plomo y destacan el potencial de MASnI_3 como una alternativa prometedora en el desarrollo de tecnologías fotovoltaicas sostenibles.

Finalmente, esta investigación aporta una base sólida para que futuros trabajos experimentales confirmen los diseños óptimos predichos y evalúen la estabilidad en condiciones reales de operación. Adicionalmente, se sugiere extender el modelo de aprendizaje de máquina incorporando nuevas variables tales como temperatura de operación, densidad de defectos o variaciones en composiciones de la perovskita para enriquecer las predicciones. La combinación de metodologías híbridas, que integren simulaciones físicas con algoritmos de inteligencia artificial, constituye un camino prometedor para acelerar el desarrollo de tecnologías fotovoltaicas de próxima generación.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

REFERENCIAS

- [1] National Renewable Energy Laboratory, “Interactive Best Research-Cell Efficiency Chart.” Accessed: Jun. 15, 2025. [Online]. Available: <https://www.nrel.gov/pv/interactive-cell-efficiency>
- [2] C. Donnelly, “Solar PV Statistics & Facts in the UK,” GreenMatch. Accessed: Jun. 15, 2025. [Online]. Available: <https://www.greenmatch.co.uk/solar-energy/solar-pv-statistics>
- [3] A. Bist, B. Pant, G. P. Ojha, J. Acharya, M. Park, and P. S. Saud, “Novel Materials in Perovskite Solar Cells: Efficiency, Stability, and Future Perspectives,” *Nanomaterials*, vol. 13, no. 11, p. 1724, May 2023, doi: 10.3390/nano13111724.
- [4] A. Yelshibay, S. D. Bukari, B. Bapayev, and M. P. Balanay, “Conducting Polymers in Solar Cells: Insights, Innovations, and Challenges,” *Organics*, vol. 5, no. 4, pp. 640–669, Dec. 2024, doi: 10.3390/org5040034.
- [5] X. Jiang *et al.*, “Isomeric diammonium passivation for perovskite–organic tandem solar cells,” *Nature*, vol. 635, no. 8040, pp. 860–866, Nov. 2024, doi: 10.1038/s41586-024-08160-y.
- [6] Ms. Stella Wang, “JinkoSolar’s Perovskite Tandem Solar Cell Based on N-type TOPCon Sets New Record with Conversion Efficiency of 33.24%,” *jinkosolar.us*, May 2024, Accessed: Jun. 15, 2025. [Online]. Available: <https://ir.jinkosolar.com/news-releases/news-release-details/jinkosolars-perovskite-tandem-solar-cell-based-n-type-topcon-0>
- [7] M. Noman, Z. Khan, and S. T. Jan, “A comprehensive review on the advancements and challenges in perovskite solar cell technology,” *RSC Adv*, vol. 14, no. 8, pp. 5085–5131, 2024, doi: 10.1039/D3RA07518D.
- [8] P. Roy, A. Ghosh, F. Barclay, A. Khare, and E. Cuce, “Perovskite Solar Cells: A Review of the Recent Advances,” *Coatings*, vol. 12, no. 8, p. 1089, Jul. 2022, doi: 10.3390/coatings12081089.
- [9] C. Rojas-Rincón, Y. Vélez-Galvis, M. Botero-Londoño, E. Reyes-Vera, and A. Sepúlveda, “Machine Learning for the Prediction of Perovskite Solar Cell Performance: A Brief Review,” 2025, pp. 15–24. doi: 10.1007/978-3-031-88995-0_2.
- [10] S. Datta, A. Baul, G. C. Sarker, P. K. Sadhu, and D. R. Hodges, “A Comprehensive Review of the Application of Machine Learning in Fabrication and Implementation of Photovoltaic Systems,” *IEEE Access*, vol. 11, pp. 77750–77778, 2023, doi: 10.1109/ACCESS.2023.3298542.
- [11] A. H. Rumman *et al.*, “Data-driven design for enhanced efficiency of Sn-based perovskite solar cells using machine learning,” *APL Machine Learning*, vol. 1, no. 4, Dec. 2023, doi: 10.1063/5.0177271.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

- [12] W. Li *et al.*, “Performance prediction and optimization of perovskite solar cells based on the Bayesian approach,” *Solar Energy*, vol. 262, p. 111853, Sep. 2023, doi: 10.1016/j.solener.2023.111853.
- [13] N. Parikh *et al.*, “Is machine learning redefining the perovskite solar cells?,” *Journal of Energy Chemistry*, vol. 66, pp. 74–90, Mar. 2022, doi: 10.1016/j.jechem.2021.07.020.
- [14] D. Burkitt, J. Searle, D. A. Worsley, and T. Watson, “Sequential Slot-Die Deposition of Perovskite Solar Cells Using Dimethylsulfoxide Lead Iodide Ink,” *Materials*, vol. 11, no. 11, p. 2106, Oct. 2018, doi: 10.3390/ma11112106.
- [15] S. Datta, A. Baul, G. C. Sarker, P. K. Sadhu, and D. R. Hodges, “A Comprehensive Review of the Application of Machine Learning in Fabrication and Implementation of Photovoltaic Systems,” *IEEE Access*, vol. 11, pp. 77750–77778, 2023, doi: 10.1109/ACCESS.2023.3298542.
- [16] N. E. I. Boukortt, C. Triolo, S. Santangelo, and S. Patané, “All-Perovskite Tandem Solar Cells: From Certified 25% and Beyond,” *Energies (Basel)*, vol. 16, no. 8, p. 3519, Apr. 2023, doi: 10.3390/en16083519.
- [17] Q. Wei, W. Zi, Z. Yang, and D. Yang, “Photoelectric performance and stability comparison of MAPbI₃ and FAPbI₃ perovskite solar cells,” *Solar Energy*, vol. 174, pp. 933–939, Nov. 2018, doi: 10.1016/j.solener.2018.09.057.
- [18] F. H. Alharbi and S. Kais, “Theoretical limits of photovoltaics efficiency and possible improvements by intuitive approaches learned from photosynthesis and quantum coherence,” *Renewable and Sustainable Energy Reviews*, vol. 43, pp. 1073–1089, Mar. 2015, doi: 10.1016/j.rser.2014.11.101.
- [19] B.-H. Wu, W.-T. Liu, T.-Y. Chen, T.-P. Perng, J.-H. Huang, and L.-J. Chen, “Plasmon-enhanced photocatalytic hydrogen production on Au/TiO₂ hybrid nanocrystal arrays,” *Nano Energy*, vol. 27, pp. 412–419, Sep. 2016, doi: 10.1016/j.nanoen.2016.07.029.
- [20] C. K. Bhaskare and R. G. Pawashe, “Separation and extractive spectrophotometric determination of palladium and platinum with 6-aminoquinoxaline-2,3-dithiol,” *Analyst*, vol. 106, no. 1266, p. 1005, 1981, doi: 10.1039/an9810601005.
- [21] A. N. Hanna, U. S. Bhansali, M. A. Khan, and H. N. Alshareef, “Characterization of current transport in ferroelectric polymer devices,” *Org Electron*, vol. 15, no. 1, pp. 22–28, Jan. 2014, doi: 10.1016/j.orgel.2013.10.009.
- [22] G. R. Lorzing, K. P. Balto, A. M. Antonio, B. A. Trump, C. M. Brown, and E. D. Bloch, “Elucidating the Structure of the Metal–Organic Framework Ru-HKUST-1,” *Chemistry of Materials*, vol. 32, no. 18, pp. 7710–7715, Sep. 2020, doi: 10.1021/acs.chemmater.0c01944.
- [23] J. Bisquert, *The Physics of Solar Cells*, vol. 1. Castellana: CRC Press is an imprint of Taylor & Francis Group, 2018.
- [24] T. G. Mayerhöfer, S. Pahlow, and J. Popp, “The Bouguer-Beer-Lambert Law: Shining Light on the Obscure,” *ChemPhysChem*, vol. 21, no. 18, pp. 2029–2046, Sep. 2020, doi: 10.1002/cphc.202000464.
- [25] P. Rafieipour, A. Mohandes, M. Moaddeli, and M. Kanani, “Integrating transfer matrix method into SCAPS-1D for addressing optical losses and per-layer optical

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

- properties in perovskite/Silicon tandem solar cells,” Aug. 2023, doi: <https://doi.org/10.48550/arXiv.2308.01132>.
- [26] J. Caicho-Caranqui *et al.*, “Towards sustainable bioplasticizers from biomass to polymers applications: A review,” *Sustainable Materials and Technologies*, vol. 43, p. e01194, Apr. 2025, doi: 10.1016/j.susmat.2024.e01194.
- [27] I. García, “Electrical Characteristics of Solar Cells,” in *Fundamentals of Solar Cells and Photovoltaic Systems Engineering*, Elsevier, 2025, pp. 99–128. doi: 10.1016/B978-0-323-96105-9.00004-5.
- [28] J. Nelson, *The Physics of Solar Cells*. Imperial College Press and Distributed by Worldscientific publishing co., 2003. doi: 10.1142/p276.
- [29] R. Kumar *et al.*, “Different energy storage techniques: recent advancements, applications, limitations, and efficient utilization of sustainable energy,” *J Therm Anal Calorim*, vol. 149, no. 5, pp. 1895–1933, Mar. 2024, doi: 10.1007/s10973-023-12831-9.
- [30] J. Nelson, *The Physics of Solar Cells*. Imperial College Press and Distributed by Worldscientific publishing co., 2003. doi: 10.1142/p276.
- [31] J. Li, Y. Wu, S. Ma, M. Chen, B. Zhang, and B. Jiang, “Analysis of photovoltaic array maximum power point tracking under uniform environment and partial shading condition: A review,” *Energy Reports*, vol. 8, pp. 13235–13252, Nov. 2022, doi: 10.1016/j.egy.2022.09.192.
- [32] M. A. Green, “Solar cell fill factors: General graph and empirical expressions,” *Solid State Electron*, vol. 24, no. 8, pp. 788–789, Aug. 1981, doi: 10.1016/0038-1101(81)90062-9.
- [33] W. Shockley and H. J. Queisser, “Detailed Balance Limit of Efficiency of *p-n* Junction Solar Cells,” *J Appl Phys*, vol. 32, no. 3, pp. 510–519, Mar. 1961, doi: 10.1063/1.1736034.
- [34] Fraunhofer ISE, “Fraunhofer Institute for Solar Energy Systems ISE.” Accessed: Jun. 15, 2025. [Online]. Available: <https://www.ise.fraunhofer.de/en/business-areas/photovoltaics-materials-cells-and-modules/perovskite-thin-film-photovoltaics.html>
- [35] X. Jia *et al.*, “The progress and challenges of tin-lead alloyed perovskites: Toward the development of large-scale all-perovskite tandem solar cells,” *Chem*, vol. 11, no. 1, p. 102384, Jan. 2025, doi: 10.1016/j.chempr.2024.12.002.
- [36] A. S. Abdulkarim *et al.*, “An overview of tin based perovskite solar cells: Stability and efficiency,” *Current Applied Physics*, vol. 71, pp. 190–198, Mar. 2025, doi: 10.1016/j.cap.2025.01.010.
- [37] “Silvaco Herramientas Propiedad intelectual Soluciones Serv.” Accessed: Jun. 15, 2025. [Online]. Available: <https://silvaco.com>
- [38] “COMSOL.” Accessed: Jun. 15, 2025. [Online]. Available: <https://www.comsol.com>
- [39] M. Bulgerman, “Simulation programme SCAPS-1D for thin film solar cells,” developed at ELIS, University of Gent. Accessed: Jun. 15, 2025. [Online]. Available: <https://scaps.elis.ugent.be>

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

- [40] J. Botero-Valencia *et al.*, “Machine Learning in Sustainable Agriculture: Systematic Review and Research Perspectives,” *Agriculture*, vol. 15, no. 4, p. 377, Feb. 2025, doi: 10.3390/agriculture15040377.
- [41] C. Hernandez, C. Castañeda, S. Castro, B. Medina, A. Sepulveda, and E. Reyes, “Caracterización de celdas solares de perovskita mediante un análisis exploratorio de datos de variables eléctricas, band gap y área,” Oct. 2024, doi: <https://doi.org/10.15665/rp.v22i2.3485>.
- [42] Y. Velez-Galvis, J. Echeverri-Perez, E. Reyes-Vera, E. Gonzalez-Valencia, B. Medina, and A. Sepulveda, “Simulation of lead-free perovskite solar cell based on MASnI₃ compared with FAPbI₃,” in *Optica Latin America Optics and Photonics Conference (LAOP) 2024*, Washington, D.C.: Optica Publishing Group, 2024, p. Tu4A.16. doi: 10.1364/LAOP.2024.Tu4A.16.
- [43] J. Soto-Perdomo, Y. R. Moreno, J. A. Moreno, J. Montoya-Cardona, E. Reyes-Vera, and J. Herrera-Ramirez, “Temperature Sensor Based on Modal Distribution in Long-Period Fiber Gratings: A Deep Learning Approach,” in *ECSA-11*, Basel Switzerland: MDPI, Nov. 2024, p. 56. doi: 10.3390/ecsa-11-20417.
- [44] S. Valencia-Garzon, J. Montoya-Cardona, E. Reyes-Vera, and E. Gonzalez-Valencia, “Inverse Design of Optical Microring Resonators Using a Supervised Machine Learning Approach,” in *Optica Latin America Optics and Photonics Conference (LAOP) 2024*, Washington, D.C.: Optica Publishing Group, 2024, p. M4A.3. doi: 10.1364/LAOP.2024.M4A.3.
- [45] M. Solarte-Sanchez, D. Marquez-Viloria, A. E. Castro-Ospina, E. Reyes-Vera, N. Guerrero-Gonzalez, and J. Botero-Valencia, “m-QAM Receiver Based on Data Stream Spectral Clustering for Optical Channels Dominated by Nonlinear Phase Noise,” *Algorithms*, vol. 17, no. 12, p. 553, Dec. 2024, doi: 10.3390/a17120553.
- [46] N. Shrivastav, J. Madan, and R. Pandey, “Predicting photovoltaic efficiency in Cs-based perovskite solar cells: A comprehensive study integrating SCAPS simulation and machine learning models,” *Solid State Commun*, vol. 380, p. 115437, Mar. 2024, doi: 10.1016/j.ssc.2024.115437.
- [47] S. Jiang *et al.*, “Machine learning (ML)-assisted optimization doping of KI in MAPbI₃ solar cells,” *Rare Metals*, vol. 40, no. 7, pp. 1698–1707, Jul. 2021, doi: 10.1007/s12598-020-01579-y.
- [48] V. Deswal and S. Baghel, “Designing and optimization of a highly efficient and new lead-free Cs₂RbGaI₆ based double perovskite solar cell through SCAPS-1D and machine learning,” *Inorg Chem Commun*, vol. 176, p. 114316, Jun. 2025, doi: 10.1016/j.inoche.2025.114316.
- [49] L. Breiman, “Bagging predictors,” *Mach Learn*, vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1023/A:1018054314350.
- [50] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009. doi: 10.1007/978-0-387-84858-7.
- [51] X. Chen, M. Wang, and H. Zhang, “The use of classification trees for bioinformatics,” *WIREs Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 55–63, Jan. 2011, doi: 10.1002/widm.14.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

- [52] S. Gey and E. Nedelec, “Model Selection for CART Regression Trees,” *IEEE Trans Inf Theory*, vol. 51, no. 2, pp. 658–670, Feb. 2005, doi: 10.1109/TIT.2004.840903.
- [53] H. Kim and W.-Y. Loh, “Classification Trees With Unbiased Multiway Splits,” *J Am Stat Assoc*, vol. 96, no. 454, pp. 589–604, Jun. 2001, doi: 10.1198/016214501753168271.
- [54] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. New York, NY: Springer US, 2021. doi: 10.1007/978-1-0716-1418-1.
- [55] L. Breiman, “Random Forest,” *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [56] N. Cardona, “Predicción y selección de variables con bosques aleatorios en presencia de variables correlacionadas,” Maestría en Ciencias - Estadística, Universidad Nacional de Colombia, Medellín, 2019.
- [57] B. Gregorutti, B. Michel, and P. Saint-Pierre, “Correlation and variable importance in random forests,” *Stat Comput*, vol. 27, no. 3, pp. 659–678, May 2017, doi: 10.1007/s11222-016-9646-1.
- [58] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” *BMC Genomics*, vol. 21, no. 1, p. 6, Dec. 2020, doi: 10.1186/s12864-019-6413-7.
- [59] C. Willmott and K. Matsuura, “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,” *Clim Res*, vol. 30, pp. 79–82, 2005, doi: 10.3354/cr030079.
- [60] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature,” *Geosci Model Dev*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, doi: 10.5194/gmd-7-1247-2014.
- [61] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr. 2002, doi: 10.1109/4235.996017.
- [62] R. T. Marler and J. S. Arora, “Survey of multi-objective optimization methods for engineering,” *Structural and Multidisciplinary Optimization*, vol. 26, no. 6, pp. 369–395, Apr. 2004, doi: 10.1007/s00158-003-0368-6.
- [63] J. A. Aranda Pinilla and J. A. Orjuela Castro, “Optimización multiobjetivo en la gestión de cadenas de suministro de biocombustibles. Una revisión de la literatura,” *Ingeniería*, vol. 20, no. 1, Mar. 2015, doi: 10.14483/udistrital.jour.reving.2015.1.a03.
- [64] K. Miettinen, *Nonlinear Multiobjective Optimization*, vol. 12. Boston, MA: Springer US, 1998. doi: 10.1007/978-1-4615-5563-6.
- [65] M. Baro and P. Borgohain, “SCAPS-1D Device Simulation of Highly Efficient Perovskite Solar Cells Using Diverse Charge Transport Layers,” *J Electron Mater*, vol. 52, no. 11, pp. 7623–7644, Nov. 2023, doi: 10.1007/s11664-023-10681-7.
- [66] S. B. Ivriq, M. H. Mohammadi, and R. S. Davidsen, “Enhancing photovoltaic efficiency in Half-Tandem MAPbI₃/MASnI₃ Perovskite solar cells with triple core-shell plasmonic nanoparticles,” *Sci Rep*, vol. 15, no. 1, p. 1478, Jan. 2025, doi: 10.1038/s41598-025-85243-4.
- [67] M. Uribe Galeano, “Obtención de dispersiones coloidales de nanopartículas de TiO₂ para su aplicación como capa transportadora de electrones en celdas solares de

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

- perovskita con estructura p-i-n,” Trabajo de grado profesional, Ingeniería de Materiales, Universidad de Antioquia, Medellín, Antioquia, Colombia, 2022. Accessed: Jun. 15, 2025. [Online]. Available: <https://hdl.handle.net/10495/30105>
- [68] J. Lee, T. Son, K. Min, S. Park, Y. Kim, and J. Seo, “Rationally designed hole transporting layer system for efficient and stable perovskite solar cells,” *EcoMat*, vol. 5, no. 11, Nov. 2023, doi: 10.1002/eom2.12414.
- [69] Md. S. Rahman, N. Ahmed, T. Paul, and Md. M. Rahman, “Performance Evaluation of Lead Free $\text{CH}_3\text{NH}_3\text{SnI}_3$ Perovskite Solar Cell: A Simulation Approach by SCAPS-1D,” in *2024 Third International Conference on Power, Control and Computing Technologies (ICPC2T)*, IEEE, Jan. 2024, pp. 392–397. doi: 10.1109/ICPC2T60072.2024.10474693.
- [70] D. Stanić *et al.*, “Simulation and Optimization of FAPbI₃ Perovskite Solar Cells with a BaTiO₃ Layer for Efficiency Enhancement,” *Materials*, vol. 15, no. 20, p. 7310, Oct. 2022, doi: 10.3390/ma15207310.
- [71] H. Li, Y. Huang, M. Zhu, P. Yan, and C. Sheng, “Analyzing Efficiency of Perovskite Solar Cells Under High Illumination Intensities by SCAPS Device Simulation,” *Nanomaterials*, vol. 15, no. 4, p. 286, Feb. 2025, doi: 10.3390/nano15040286.
- [72] A.-G. Ulăreanu and A. Dragulinescu, “SCAPS-1D simulation and optimization of an organic solar cell,” in *Advances in 3OM: Opto-Mechatronics, Opto-Mechanics, and Optical Metrology (3OM 2023)*, M. Guina, J. P. Rolland-Thompson, V.-F. Duma, A. G. H. Podoleanu, and C. Sinescu, Eds., SPIE, Jun. 2024, p. 68. doi: 10.1117/12.3021754.
- [73] J. A. P. Echeverri, “DATASET_MASnI₃_FAPbI₃-PCE_FF,” Jun. 11, 2025, *OSF, Medellín*. doi: 10.17605/OSF.IO/97F84.
- [74] S. Karthick, S. Velumani, and J. Bouclé, “Experimental and SCAPS simulated formamidinium perovskite solar cells: A comparison of device performance,” *Solar Energy*, vol. 205, pp. 349–357, Jul. 2020, doi: 10.1016/j.solener.2020.05.041.
- [75] Touré Sékou, Amal Bouich, Youssouf Doumbia, and Bernabé Marí Soucase, “Comparative Analysis Using the SCAPS-1D Software on the Stability and Toxicity of the Perovskites FAPbI₃ and FASnI₃,” *Advanced Energy Conversion Materials*, pp. 308–320, Oct. 2024, doi: 10.37256/aecm.5220245549.
- [76] M. R. A. Elsayed, A. M. Elseman, A. A. Abdelmageed, H. M. Hashem, and A. Hassen, “Synthesis and numerical simulation of formamidinium-based perovskite solar cells: a predictable device performance at NIS-Egypt,” *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-37018-y.
- [77] I. Bouderbala and N.-E. Houda, “Effects of absorber layer thickness and doping density on the performance of perovskite solar cells: a simulation analysis using SCAPS-1D software,” *Algerian Journal of Engineering and Technology*, Jun. 2023, doi: <https://doi.org/10.57056/ajet.v8i1.101>.
- [78] P. Saha, S. Singh, and S. Bhattacharya, “Performance optimization of MASnI₃ perovskite solar cells: Insights into device architecture,” *Micro and Nanostructures*, vol. 191, p. 207827, Jul. 2024, doi: 10.1016/j.micrna.2024.207827.

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

- [79] A. K. Singh, S. Srivastava, A. Mahapatra, J. K. Baral, and B. Pradhan, “Performance optimization of lead free-MASnI3 based solar cell with 27% efficiency by numerical simulation,” *Opt Mater (Amst)*, vol. 117, p. 111193, Jul. 2021, doi: 10.1016/j.optmat.2021.111193.

 Institución Universitaria	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02- 2020

ANEXOS

	INFORME FINAL TRABAJO DE GRADO	Código	FDE 089
		Versión	04
		Fecha	24-02-2020

FIRMA ESTUDIANTES _____ FIRMA ASESORES _____ <div style="margin-left: 300px;"> <u>Erick Reyes Vera</u> <u>ESTEBAN GONZALEZ</u> </div>
FECHA ENTREGA: <u>16 de junio 2025</u>