Generación de láminas funiculares guiada por física e incertidumbre

Un surrogado eficiente con incertidumbre y un guiado adaptativo para el pipeline Hybrid PI-DDPM

Daniel Ariza García
dirigido por
Diego Canales Aguilera

Trabajo Fin de Máster
Máster Universitario en Inteligencia Artificial

Universidad Loyola Andalucía
Curso: 2025 – 2026

Generación de láminas funiculares guiada por física e incertidumbre

Daniel Ariza García

Dirigido por: Diego Canales Aguilera

Resumen

Las láminas funiculares son estructuras de gran eficiencia mecánica cuya forma canaliza las cargas mediante esfuerzos de membrana, pero diseñarlas es costoso, ya que la vía clásica mantiene un solver de elementos finitos dentro del bucle de optimización y devuelve una única solución por caso. Los modelos generativos ofrecen diversidad a bajo coste, mas carecen de cualquier noción de la física, y verificar cada candidato con una simulación completa devolvería el coste que se pretendía evitar. Este Trabajo Fin de Máster propone un pipeline modular de generación de láminas guiada por física, formado por dos redes que se entrenan por separado y solo cooperan durante la inferencia. El Arquitecto es un prior de difusión que aprende el repertorio de formas plausibles, y el Ingeniero un surrogado mecánico diferenciable cuyo gradiente corrige la trayectoria de muestreo hacia geometrías más eficientes. Esta separación evita la optimización conjunta y el equilibrio adversario, e inyecta el conocimiento físico sin reentrenar el prior. Sobre esa base, el Ingeniero se entrena directamente sobre estados ruidosos, lo que aborda la brecha de Jensen que aparecería al evaluar una física no lineal sobre estimaciones limpias. Se propone además un surrogado espectral con incertidumbre, EquiNO+σ^2, que predice el factor de membrana con mayor precisión que el baseline convolucional empleando del orden de 3,5 veces menos parámetros, y un guiado adaptativo cuya forma temporal se deriva de la incertidumbre calibrada del propio surrogado, sin hiperparámetros ajustados a mano. Un análisis basado en la relación señal-ruido espectral y en la brecha de Jensen explica por qué el error del surrogado en la ventana de guiado es varianza irreducible. Los experimentos confirman que el guiado eleva sustancialmente la eficiencia mecánica de las láminas generadas, y que el guiado adaptativo iguala a una campana ajustada a mano con una diversidad ligeramente mayor.

Palabras clave: láminas funiculares, modelos de difusión, guiado físico, surrogados neuronales, cuantificación de incertidumbre, factor de membrana.

Abstract

Funicular shells are highly efficient structures whose shape channels loads through membrane action, yet designing them is costly, since the classical route keeps a finite-element solver inside the optimisation loop and returns a single solution per case. Generative models offer diversity at low cost but have no notion of physics, and verifying each candidate with a full simulation would restore the very cost one sought to avoid. This Master's Thesis proposes a modular pipeline for physics-guided shell generation, built from two networks trained separately and coupled only at inference. The Architect is a diffusion prior that learns the repertoire of plausible shapes, and the Engineer a differentiable mechanical surrogate whose gradient steers the sampling trajectory towards more efficient geometries. This separation avoids joint optimisation and adversarial balancing, and injects physical knowledge without retraining the prior. On this basis, the Engineer is trained directly on noisy states, which addresses the Jensen gap that would arise from evaluating non-linear physics on clean estimates. A spectral surrogate with uncertainty, EquiNO+σ^2, is further proposed, predicting the membrane factor more accurately than the convolutional baseline while using about 3,5 times fewer parameters, together with an adaptive guidance whose temporal shape is derived from the surrogate's own calibrated uncertainty, without hand-tuned hyperparameters. An analysis based on the spectral signal-to-noise ratio and on the Jensen gap explains why the surrogate error within the guidance window is irreducible variance. Experiments confirm that guidance substantially raises the mechanical efficiency of the generated shells, and that adaptive guidance matches a hand-tuned bell schedule with slightly greater diversity.

Keywords: funicular shells, diffusion models, physics guidance, neural surrogates, uncertainty quantification, membrane factor.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

1. Contexto y motivación

Las estructuras laminares, conocidas como *shells*, son superficies curvas de pequeño espesor que constituyen una de las familias más elegantes y eficientes de la ingeniería estructural. Su repertorio de formas es amplio, desde cúpulas y bóvedas hasta paraboloides hiperbólicos y láminas de forma libre, y todas comparten un mismo principio, y es que cuando la geometría está alineada con el recorrido de las fuerzas internas, las cargas se transmiten mediante esfuerzos de membrana contenidos en el propio plano de la lámina, la flexión queda reducida a una contribución residual y la estructura puede cubrir grandes áreas con muy poco material. Una geometría que trabaja en este régimen se denomina **funicular**, y encontrarla es, en esencia, un problema de forma.

La idea tiene una historia brillante, de los modelos colgantes con los que Gaudí buscaba las formas de equilibrio (Huerta, 2006) a las más de ochocientas láminas de hormigón de Félix Candela (Faber, 1963), y vive hoy un renacimiento de la mano de la fabricación digital. El puente Striatus de Venecia, construido por el Block Research Group de la ETH de Zúrich junto a Zaha Hadid Architects, es una lámina de bloques de hormigón impresos en 3D que se sostiene sin armadura alguna, solo por su forma (Bhooshan y cols., 2022). La Figura 1.1 recoge ambos mundos, la herencia de Candela y la fabricación digital actual, unidos por el mismo principio estructural. La impresión 3D de hormigón refuerza además el interés por la funicularidad, porque las mezclas imprimibles apenas resisten tracciones y una geometría que trabaje a compresión no es una elegancia, sino una condición de

viabilidad constructiva (Wangler, Roussel, Bos, Salet, y Flatt, 2019).

Figura 1.1: Láminas funiculares, ayer y hoy. Izquierda: el restaurante del Oceanogràfic de Valencia, con los paraboloides hiperbólicos característicos de Félix Candela, uno de sus últimos diseños. Derecha: el puente Striatus en Venecia (Block Research Group y Zaha Hadid Architects), impreso en 3D y sin armadura. En ambos casos la eficiencia estructural procede de la forma.

Diseñar estas estructuras, sin embargo, es mucho un problema difícil de resorver. Precisamente porque el espesor es minúsculo frente al ancho y al alto de la lámina, su comportamiento es extremadamente sensible a la forma ya que pequeñas desviaciones de la geometría funicular despiertan flexiones que el material apenas puede absorber. Simular ese comportamiento en hormigón exige modelos de lámina delgada resueltos por elementos finitos, un proceso computacionalmente costoso en sí mismo.

Además la vía clásica para automatizar el diseño, la optimización topológica, no soluciona el problema debido a que el solver de elementos finitos entra dentro del bucle de optimización y debe ejecutarse en cada iteración, lo que convierte el procedimiento en algo tedioso y lento, y además devuelve una única figura óptima para cada combinación de cargas y apoyos, no una variedad de alternativas. Para una fase conceptual de diseño, en la que se necesita comparar familias de candidatos, este planteamiento se queda corto por coste y por falta de diversidad.

La inteligencia artificial generativa aparece como una solución natural, y desde luego está en su momento de mayor auge. En lugar de buscar el óptimo de un problema, un modelo generativo aprende una distribución de diseños válidos y muestrea candidatos diversos en fracciones de segundo. Dentro de esta familia, los modelos de difusión (Ho, Jain, y Abbeel, 2020) han sustituido a las alternativas adversariales por su entrenamiento estable y su cobertura de la distribución (Dhariwal

y Nichol, 2021), sustentan los grandes generadores de imagen actuales y, para este trabajo, tienen una propiedad decisiva puesto que generan a través de una trayectoria de estados intermedios que puede intervenirse paso a paso.

En cambio, **un modelo de difusión no tiene ninguna capacidad de entender la física**, que es una carencia fundamental clave de esta memoria. Aprende la apariencia de sus datos de entrenamiento, no las leyes que los hacen válidos de modo que entrenado con geometrías de lámina producirá superficies de aspecto convincente, pero nada en su interior le impide generar formas que, sometidas a cargas, desarrollan flexiones inadmisibles. Por tanto, verificar cada candidato con una simulación completa devolvería al bucle justo el coste del que se quería escapar.

Aquí es donde cobran importancia los **modelos surrogados** basados en redes neuronales que aprenden a imitar al simulador y devuelven la respuesta física en milisegundos, de forma además diferenciable, lo que permite usarlas no solo para evaluar sino para **corregir** la generación mediante gradientes. Se trata de una línea de investigación plenamente internacional. El grupo de George Karniadakis en la Universidad de Brown introdujo DeepONet para aprender operadores entre funciones (Lu, Jin, Pang, Zhang, y Karniadakis, 2021), el de Anima Anandkumar en Caltech propuso el Fourier Neural Operator, que parametriza el operador en el dominio de la frecuencia (Li y cols., 2021), el laboratorio de Faez Ahmed en el MIT ha llevado los generadores guiados por rendimiento al diseño de ingeniería (Mazé y Ahmed, 2023; Regenwetter, Nobari, y Ahmed, 2022) y el grupo de Dennis Kochmann en la ETH de Zúrich ha incorporado los residuos de la física al propio entrenamiento de los modelos de difusión (Bastek, Sun, y Kochmann, 2024). En España, esta línea tiene uno de sus focos más activos en la Universidad de Zaragoza, donde el grupo de Elías Cueto, en colaboración con el de Francisco Chinesta en el ENSAM de París, ha impulsado desde los gemelos híbridos (Chinesta, Cueto, Abisset-Chavanne, Duval, y El Khaldi, 2020) hasta redes neuronales que incorporan la estructura matemática o las leyes de la termodinámica del problema como sesgo inductivo (Cueto y Chinesta, 2023; Hernández, Badías, González, Chinesta, y Cueto, 2021), demostrando que un surrogado rinde más y es más fiable cuanto más conocimiento físico incorpora su diseño. De ese mismo entorno procede el precedente directo de este trabajo y su conjunto de datos (Lourenço y cols., 2024).

Este TFM nace exactamente de la idea de que si los generadores de difusión producen formas diversas

pero ciegas a la mecánica, y los modelos surrogados ofrecen una física rápida y diferenciable, podrían combinarse en un pipeline que genere láminas funiculares que no solo parezcan plausibles, sino que sean físicamente coherentes y mecánicamente eficientes.

Materializar esa idea plantea tres retos encadenados. El primero es generar geometrías plausibles y diversas sin un solver en el bucle. El segundo, evaluar su comportamiento mecánico de forma rápida y diferenciable durante la propia generación. Y el tercero, el más sutil, decidir **cuándo** conviene fiarse de esa evaluación, porque durante buena parte de la trayectoria de difusión la geometría está tan contaminada de ruido que cualquier juicio mecánico es necesariamente incierto.

La respuesta de este trabajo es el pipeline Hybrid PI-DDPM, formado por dos redes especializadas que se entrenan por separado y solo cooperan al generar. El **Arquitecto** es un modelo de difusión que aprende el repertorio de formas de lámina plausibles y aporta la diversidad. El **Ingeniero** es un surrogado mecánico diferenciable que estima la respuesta estructural y, a partir de ella, el factor de membrana, la métrica que resume cuánta de la energía se conduce por membrana frente a flexión, y su gradiente corrige la trayectoria de muestreo hacia geometrías más eficientes.

Dos decisiones de diseño distinguen a este pipeline. La primera es la separación modular, ya que al no entrenar un único modelo que genere y respete la física a la vez se evitan las optimizaciones conjuntas frágiles y los equilibrios adversarios, y el conocimiento físico se inyecta en inferencia sin reentrenar nada. La segunda es que el Ingeniero es **consciente del ruido** ya que se entrena directamente sobre los estados ruidosos que recorrerá el muestreo, evitando el sesgo sistemático (la brecha de Jensen) que aparece al evaluar física no lineal sobre estimaciones limpias. Sobre esa base, la corrección física se concentra en el ruido intermedio, donde la geometría es a la vez informativa para el Ingeniero y todavía maleable bajo el Arquitecto.

De ahí resultan las cuatro aportaciones de la memoria. La primera es el propio pipeline modular de difusión guiada por física. La segunda es un Ingeniero espectral más eficiente y con incertidumbre (EquiNO+σ^2), que predice mejor el factor de membrana que el baseline convolucional con del orden de 3,5 veces menos parámetros. La tercera es un guiado adaptativo cuya forma temporal se deriva de la incertidumbre del propio surrogado, sin hiperparámetros ajustados a mano. Y la cuarta es un análisis teórico del error, mediante la relación señal-ruido espectral y la brecha de Jensen, que

explica por qué todo lo anterior funciona. En todo el estudio el Arquitecto se mantiene congelado y reproducible, de modo que cualquier diferencia medida se atribuye al Ingeniero o al guiado, que son las piezas investigadas.

Parte de este trabajo, en concreto los fundamentos del Arquitecto y del Ingeniero sobre los que se apoya el resto de la memoria, se ha presentado en un congreso internacional, el *Congress on Numerical Methods in Engineering* (CMN 2026) organizado por la SEMNI y la APMTAC y celebrado en Gijón en julio de 2026, y se ha enviado a una revista indexada en el primer cuartil (Q1) del Journal Citation Reports.

2. Objetivos

El objetivo general de este TFM es desarrollar y evaluar un pipeline modular de generación de láminas funiculares guiada por física, basado en un prior de difusión y un surrogado mecánico consciente del ruido, y profundizar en el surrogado y la regla de guiado, dotando a esta última de una forma derivada de la incertidumbre del propio modelo en lugar de un ajuste manual.

Este objetivo general se concreta en los siguientes objetivos específicos:

- Formalizar una representación común de la geometría, la carga y la respuesta mecánica de las láminas, adecuada para arquitecturas neuronales sobre rejilla, y establecer el marco teórico que conecta difusión, guiado físico, brecha de Jensen e incertidumbre.
- Entrenar un prior de difusión estable sobre mapas de altura de láminas, en los regímenes con y sin hueco, y congelarlo como generador fijo y reproducible del estudio.
- Construir y comparar surrogados mecánicos conscientes del ruido de distintas familias (convolucional y espectral), midiendo su precisión de campo y de factor de membrana a lo largo de todo el espectro de ruido de la difusión.
- Explicar el comportamiento observado a partir del régimen de información que impone la difusión, en el que la relación señal-ruido cae modo a modo, la varianza posterior se vuelve irreducible y persiste una brecha de Jensen residual.
- Dotar al surrogado de una incertidumbre heterocedástica, evaluar su calibración y transfor-

marla en un schedule temporal de guiado sin hiperparámetros de forma.

- Evaluar la generación guiada de forma rigurosa a través de la calidad mecánica mediante el factor de membrana, diversidad mediante la cobertura del espacio de diseño, efecto de la fuerza de guiado y significancia estadística de las comparaciones.

3. Estructura de la memoria

El resto del documento sigue el hilo natural del trabajo. El Capítulo 2 sitúa la propuesta frente al *form-finding* estructural, los modelos generativos, el guiado de difusión y los modelos surrogados neuronales, delimitando con precisión qué es nuevo. El Capítulo 3 desarrolla el marco matemático explicando los conceptos de difusión, guiado, la brecha de Jensen, el régimen de información por relación señal-ruido y la incertidumbre calibrada. El Capítulo 4 describe los datos, las arquitecturas de surrogado comparadas, su entrenamiento, el guiado y el protocolo de evaluación. El Capítulo 5 presenta los resultados, del punto de partida a las aportaciones, con su discusión y sus límites. Finalmente, el Capítulo 6 recoge las conclusiones y las líneas de trabajo futuro.

Capítulo 2

Estado del arte

Este capítulo sitúa el trabajo dentro de la literatura, pero no pretende ser una revisión exhaustiva, sino acotar las ideas que motivan directamente el diseño del sistema. Son cuatro. La primera es el coste de los métodos clásicos de *form-finding* que mantienen un solver de elementos finitos en el bucle. La segunda, las limitaciones de los generadores basados en VAE y GAN frente a la alternativa de los modelos de difusión. La tercera, el papel de los surrogados neuronales como sustitutos del solver y la forma en que se inyecta la física en el guiado. Y la cuarta, la cuantificación de la incertidumbre, cuyo uso como señal para decidir el guiado es el *gap* que esta memoria cubre.

1. Form-finding y el coste del solver

En las láminas funiculares la forma es, en sí misma, la principal variable de diseño mecánico puesto que una geometría bien elegida conduce las cargas hasta los apoyos mediante esfuerzos de membrana y mantiene la flexión en niveles residuales, lo que permite cubrir grandes áreas con muy poco material (Adriaenssens, Block, Veenendaal, y Williams, 2014; Block y Ochsendorf, 2007; Isler, 1961). Encontrar esas formas de manera automática se ha abordado tradicionalmente con métodos de *form-finding* y con optimización topológica, que distribuyen material sobre un dominio imponiendo equilibrio o minimizando una energía mecánica bajo restricciones (Bendsøe y Sigmund, 2003; Sigmund y Maute, 2013).

Estos métodos funcionan y están bien fundamentados, pero arrastran dos limitaciones para una fase

conceptual de diseño. La primera es el coste de cada iteración que requiere resolver una simulación por elementos finitos completa, cuyo coste crece de forma desfavorable con la resolución de la malla (Aage, Andreassen, Lazarov, y Sigmund, 2017), de modo que el solver permanece dentro del bucle de optimización. La segunda es de fondo al tratarse de procedimientos de diseño puntuales, que devuelven una única solución para una combinación fija de cargas, apoyos y dominio, y cualquier cambio obliga a reiniciar el proceso.

En fases tempranas se necesita justo lo contrario, generar y comparar familias diversas de candidatos a bajo coste, y es ahí donde aparecen los modelos generativos. Estos modelos desplazan el coste al entrenamiento y permiten muestrear candidatos en fracciones de segundo, dejando el solver fuera del bucle de generación (Regenwetter y cols., 2022).

2. Modelos generativos: de VAE y GAN a la difusión

Entre las primeras familias de modelos generativos profundos aplicadas al diseño se encuentran los autoencoders variacionales, o VAE (Kingma y Welling, 2014), y las redes generativas adversarias, o GAN (Goodfellow y cols., 2014). Los VAE aprenden una representación latente continua y regularizada desde la que es posible reconstruir y generar nuevas muestras. Sin embargo, su función de entrenamiento establece un compromiso entre la fidelidad de la reconstrucción y la organización probabilística del espacio latente. Cuando se emplean pérdidas elemento a elemento, el decodificador puede aproximar el promedio de varias soluciones plausibles, lo que tiende a suavizar bordes, discontinuidades y detalles geométricos finos (Larsen, Sønderby, Larochelle, y Winther, 2016). En diseño estructural, esta pérdida de definición puede alterar elementos delgados, contornos de huecos o uniones entre componentes, aunque la geometría obtenida mantenga una apariencia global razonable.

Además, una regularización excesiva o un decodificador demasiado expresivo pueden provocar el colapso del posterior. En este caso, el modelo aprende a ignorar total o parcialmente las variables latentes, reduciendo la información disponible para controlar las geometrías generadas (He, Spokoyny, Neubig, y Berg-Kirkpatrick, 2019). El problema no se limita, por tanto, a la nitidez de las muestras, sino que también puede afectar a la capacidad del espacio latente para representar familias estructurales diferentes y permitir una exploración útil del diseño.

Las GAN suelen producir geometrías más nítidas porque sustituyen la reconstrucción explícita por un entrenamiento adversario entre un generador y un discriminador (Goodfellow y cols., 2014). No obstante, este proceso requiere mantener un equilibrio delicado entre ambas redes. Si una de ellas aprende más rápido que la otra, los gradientes pueden volverse poco informativos y el entrenamiento puede oscilar, no converger o ser muy sensible a la arquitectura y a los hiperparámetros (Arjovsky, Chintala, y Bottou, 2017). También pueden sufrir colapso de modos, situación en la que el generador descubre unas pocas familias de diseños capaces de engañar al discriminador y deja de representar otras regiones de la distribución (Metz, Poole, Pfau, y Sohl-Dickstein, 2017). Este fallo resulta especialmente problemático en diseño estructural, donde el objetivo no es producir únicamente soluciones plausibles y de alta calidad, sino explorar alternativas geométricas y topológicas diversas. Además, la pérdida de diversidad puede pasar inadvertida al observar muestras individuales, ya que las geometrías generadas parecen válidas aunque el conjunto cubra solo una parte reducida del espacio de diseño.

El precedente más próximo en láminas funiculares es el trabajo de Lourenço et al. (Lourenço y cols., 2024), que combina una GAN convolucional con un surrogado mecánico de ramas paralelas y emplea el factor de membrana como medida de calidad estructural. Ese trabajo es importante porque conecta la generación geométrica con una métrica mecánica diferenciable y porque aporta el conjunto de datos de láminas que se utiliza también en esta memoria. La limitación que deja abierta no está en la métrica, sino en el modelo generativo de base, ya que como se ha mencionado, el marco adversario dificulta controlar la diversidad y modular la física durante la inferencia.

Los modelos de difusión ofrecen una alternativa más estable (Ho y cols., 2020; Sohl-Dickstein, Weiss, Maheswaranathan, y Ganguli, 2015). En lugar de un equilibrio adversario, aprenden a invertir un proceso que destruye progresivamente la señal mediante ruido gaussiano, lo que se traduce en un entrenamiento estable, una buena cobertura de la distribución (Dhariwal y Nichol, 2021) y, sobre todo, en una propiedad decisiva para este trabajo, gracias a que el generador no produce la muestra de golpe, sino que atraviesa una trayectoria de estados intermedios que puede intervenirse paso a paso.

La difusión ya se ha aplicado con éxito a la síntesis de componentes estructurales, también sobre espacios latentes (Herron y cols., 2023). La lectura *score-based* de Song et al. (Y. Song y cols., 2021)

formaliza la intervención de la trayectoria como la suma de un término generativo y un término de guiado, y es la que sostiene el guiado físico de esta memoria.

3. Surrogados neuronales y guiado físico

Como ya se ha comentado, la física ha de evaluarse durante la generación, llamar a un solver en cada paso es inviable. Una alternativa es entrenar un surrogado neuronal que aproxime el operador mecánico, de forma que dada una geometría y una carga, predice los campos de respuesta. Las arquitecturas convolucionales son la opción natural cuando las entradas y salidas viven en rejillas, y una U-Net (Ronneberger, Fischer, y Brox, 2015) combina contexto global y detalle local mediante conexiones de salto. Más allá de los mapas imagen a imagen, muchos problemas de mecánica son operadores entre funciones, lo que ha dado lugar a los *neural operators*, como DeepONet (Lu y cols., 2021) o el Fourier Neural Operator (FNO) (Li y cols., 2021), que parametrizan parte del operador en el dominio de la frecuencia y resultan eficientes cuando la información útil se concentra en pocos modos.

Inyectar física en un modelo de difusión se apoya en el guiado del muestreo. En visión por computador esto aparece como *classifier* y *classifier-free guidance* (Dhariwal y Nichol, 2021; Ho y Salimans, 2022), mientras que en problemas inversos y físicos la maniobra recurrente consiste en estimar la geometría limpia a partir del estado ruidoso, mediante la fórmula de Tweedie, y evaluar allí la condición externa (Chung, Kim, McCann, Klasky, y Ye, 2023; J. Song, Vahdat, Mardani, y Kautz, 2023). Esta estrategia es razonable para operadores lineales, pero se vuelve delicada cuando la física es no lineal puesto que evaluar el operador sobre la media posterior no equivale a la respuesta media, una discrepancia que el Marco Teórico desarrolla como brecha de Jensen.

La línea de difusión con conocimiento físico crece deprisa y se ha propuesto imponer residuos de la ecuación en derivadas parciales directamente en el entrenamiento del modelo de difusión (Bastek y cols., 2024), resolver ecuaciones bajo observación parcial guiando el muestreo con la propia ecuación (Huang, Yang, Wang, y Park, 2024) y empujar las muestras hacia el conjunto físicamente consistente durante la generación (Jacobsen, Zhuang, y Duraisamy, 2023). Todas estas variantes comparten la idea de acoplar la física al proceso de muestreo, pero ninguna se pregunta cómo varía la **fiabilidad** de esa señal física a lo largo del ruido, que es el foco de esta memoria. En optimización topológica,

TopoDiff es un precedente directo de difusión guiada por un regresor de rendimiento estructural (Giannone y Ahmed, 2023; Mazé y Ahmed, 2023), lo que obliga a matizar la novedad es que la idea general de difusión más regresor ya existe, y la diferencia de este trabajo está en el dominio, en entrenar el surrogado sobre estados ruidosos y en cómo se decide el guiado.

4. Incertidumbre y el gap de este trabajo

Si un surrogado va a guiar un generador, su error no solo afecta a una métrica posterior, sino que modifica la propia trayectoria de muestreo, de modo que conocer su incertidumbre es tan importante como conocer su error medio. La literatura distingue entre incertidumbre epistémica, que refleja desconocimiento del modelo y se reduce con más datos o con ensembles, e incertidumbre aleatoria, que describe variabilidad inherente (Kendall y Gal, 2017). Una vía práctica para estimar esta última es la regresión heterocedástica, en la que la red predice a la vez la media y la varianza dependientes de la entrada. Predecir una varianza no garantiza, sin embargo, que esté calibrada, y la calibración de regresión se mide con diagnósticos como el error de calibración esperado normalizado (Levi, Gispan, Giladi, y Fetaya, 2019).

El uso de la incertidumbre como diagnóstico está bien establecido, siendo usada en la mayoría de los trabajos para saber cuándo desconfiar de una predicción. Lo que apenas se ha explorado es incorporarla como una variable que decide el propio proceso generativo. Existen ya schedules temporales de guiado en difusión en el ámbito del *classifier-free guidance*, donde se ha propuesto modular el peso de guiado con curvas temporales paramétricas (Malarz, Kasymov, Zięba, Tabor, y Spurek, 2025) o ajustarlo paso a paso con evaluadores externos de alineamiento (Papalampidi y cols., 2025). La novedad de esta memoria no es, por tanto, hacer variar el guiado en el tiempo, sino **de dónde** sale esa variación y cómo la forma del schedule no se impone con una curva paramétrica ni la dicta un evaluador perceptual, sino que se deriva de la incertidumbre calibrada de un surrogado mecánico consciente del ruido, es decir, de una estimación de la fiabilidad física de la propia señal que guía.

Capítulo 3

Marco teórico

Este capítulo reúne las herramientas matemáticas necesarias para el resto de la memoria, siguiendo un único hilo. El método genera una geometría de lámina mediante difusión, evalúa su respuesta mecánica con un surrogado diferenciable y modifica la trayectoria de muestreo con el gradiente de un objetivo estructural. La pregunta de fondo no es cómo generar geometría plausible, sino qué información puede extraer el surrogado de un estado ruidoso, cómo aparece la incertidumbre de esa estimación, y cómo esa incertidumbre permite pasar de un guiado con forma fija a uno adaptativo. A lo largo del capítulo, un mismo objeto reaparece una y otra vez, la **varianza posterior** de la respuesta dado el estado ruidoso, y es lo que conecta el guiado, la elección de arquitectura y la incertidumbre.

1. Láminas funiculares y factor de membrana

1.1. Acción de membrana y flexión

Una lámina funicular es una superficie delgada cuya forma canaliza las cargas mediante esfuerzos contenidos en su plano tangente. En ese régimen la flexión es residual y el material trabaja de forma casi uniforme, lo que permite cubrir grandes luces con espesores mínimos (Adriaenssens y cols., 2014; Block y Ochsendorf, 2007; Isler, 1961). Desde la teoría de láminas, la respuesta combina deformaciones y esfuerzos de membrana con curvaturas y momentos flectores (Ciarlet, 2000; Reddy,

2007). En este trabajo se codifica con 13 canales por nodo,

$$y = [\, u_z;\, \varepsilon_{11}, \varepsilon_{22}, \varepsilon_{12}, n_{11}, n_{22}, n_{12};\, \kappa_{11}, \kappa_{22}, \kappa_{12}, m_{11}, m_{22}, m_{12} \,], \tag{3.1}$$

donde u_z es el desplazamiento vertical, ε_{ij} y n_{ij} son deformaciones y esfuerzos de membrana, y κ_{ij} y m_{ij} son variaciones de curvatura y momentos flectores. La separación en bloques de desplazamiento, membrana y flexión refleja comportamientos espaciales distintos, como ilustra la Figura 3.1, los campos de membrana son suaves y de baja frecuencia, mientras que los momentos flectores se concentran cerca de los apoyos y en las transiciones de curvatura, con una estructura espacial más localizada y de mayor frecuencia. Esta diferencia de contenido espectral entre membrana y flexión será clave en el resto del capítulo.

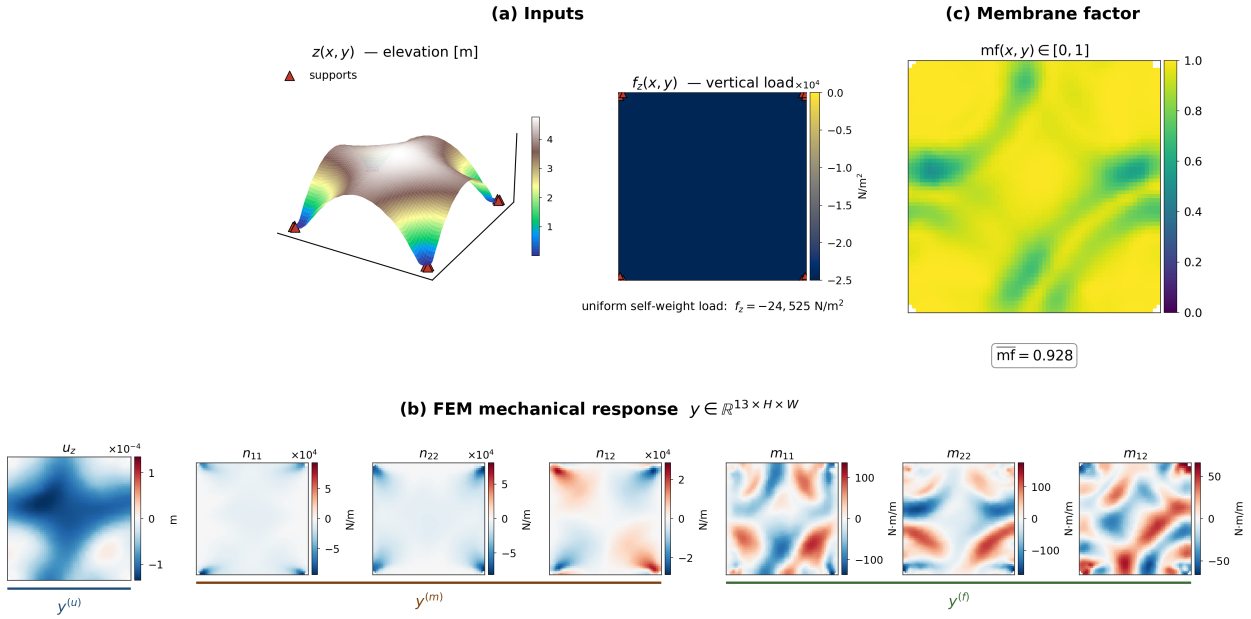

Figura 3.1: Una muestra real del conjunto de datos. (a) Entradas: el mapa de alturas $z(x, y)$, mostrado en 2.5D con sus apoyos en las esquinas (*supports*), y el campo de carga vertical de peso propio $f_z(x, y)$ (*vertical load*). (b) Respuesta mecánica de elementos finitos, $y \in \mathbb{R}^{13\times H\times W}$, separada en los tres bloques: desplazamiento u_z, esfuerzos de membrana n_{ij} (suaves, de baja frecuencia) y momentos flectores m_{ij} (localizados cerca de apoyos y transiciones de curvatura). (c) Mapa de factor de membrana $\mathrm{mf}(x, y)$, próximo a uno donde domina la membrana. Los rótulos internos están en inglés (*Inputs*, entradas; *Membrane factor*, factor de membrana; *FEM mechanical response*, respuesta mecánica por elementos finitos; *elevation*, altura).

1.2. Definición del factor de membrana

El factor de membrana mide qué fracción de la energía interna local corresponde a acción de membrana. En cada nodo se definen las densidades

$$w_m = n_{11}\varepsilon_{11} + n_{22}\varepsilon_{22} + 2n_{12}\varepsilon_{12}, \qquad w_f = m_{11}\kappa_{11} + m_{22}\kappa_{22} + 2m_{12}\kappa_{12}, \tag{3.2}$$

y el mapa de factor de membrana es

$$\mathrm{mf} = \mathrm{clip}_{[0,1]}\left(\frac{w_m}{w_m + w_f + \epsilon}\right). \tag{3.3}$$

El recorte a $[0, 1]$ es necesario porque, sobre campos predichos por una red, las densidades pueden tomar valores locales no físicos. Valores próximos a 1 indican respuesta de membrana y valores bajos indican más flexión, como muestra el panel (c) de la Figura 3.1 sobre una lámina real.

La ecuación (3.3) cumple dos papeles. Es una métrica estructural interpretable y es un **cociente no lineal** de magnitudes mecánicas. Esa no linealidad será, más adelante, la fuente de la brecha de Jensen al evaluar el objetivo sobre respuestas condicionadas a estados ruidosos.

2. Notación y planteamiento probabilístico

Con la mecánica de las láminas ya presentada, se fija la notación que usará el resto del capítulo. Sea $x_0 \in \mathbb{R}^{1\times H\times W}$ el mapa de alturas limpio de una lámina, con $H = W = 64$. El proceso de difusión produce estados ruidosos x_t para $t \in \{0, \ldots, T-1\}$, con $T = 1000$, y el coeficiente de retención de señal $\bar{\alpha}_t \in (0, 1)$ se introduce en la Sección 3. La carga vertical se denota por f_z y la respuesta mecánica de referencia por

$$y = R(x_0, f_z) \in \mathbb{R}^{13\times H\times W}, \tag{3.4}$$

donde R es el operador mecánico obtenido mediante simulación de elementos finitos y y recoge los trece canales definidos en la sección anterior. El surrogado mecánico (el **Ingeniero**) es una red diferenciable g_ϕ que recibe el estado ruidoso, el nivel de ruido y la carga, y predice la respuesta,

$$\hat{y} = g_\phi(x_t, f_z, t). \tag{3.5}$$

A partir de la respuesta se recupera el mapa de factor de membrana mf de la ecuación (3.3) y se define un objetivo escalar por muestra,

$$J(x_t) = \left(1 - \overline{\mathrm{mf}}\big(g_\phi(x_t, f_z, t)\big)\right)^2, \tag{3.6}$$

con $\overline{\mathrm{mf}}$ la media espacial del mapa, quedando el promedio sobre el lote como un detalle de implementación. Durante el guiado, el factor de membrana se calcula sobre la predicción $\hat{y}$, no sobre la respuesta FEM real. La incertidumbre del surrogado se representa mediante una varianza heteroscedástica $\sigma_\phi^2(x_t, t)$ o, equivalentemente, la log-varianza $\ell_\phi = \log \sigma_\phi^2$. Se reserva la letra w para los pesos temporales de guiado, evitando confundir incertidumbre y schedule.

3. Modelos de difusión

3.1. Fundamentos de los modelos de difusión

Un modelo de difusión genera nuevas muestras aprendiendo a invertir un proceso gradual de corrupción por ruido gaussiano (Ho y cols., 2020; Sohl-Dickstein y cols., 2015). El método consta de dos procesos complementarios. El proceso directo transforma una geometría limpia x_0 en ruido mediante una cadena de transiciones conocidas y no entrenables. El proceso inverso parte de ese ruido y utiliza una red neuronal para reconstruir progresivamente una muestra compatible con la distribución de geometrías del conjunto de datos.

La Figura 3.2 resume ambos procesos. Durante la difusión directa, la contribución de la geometría disminuye mientras aumenta la del ruido. Durante la generación, la red recorre la secuencia en sentido contrario y elimina parte del ruido en cada paso.

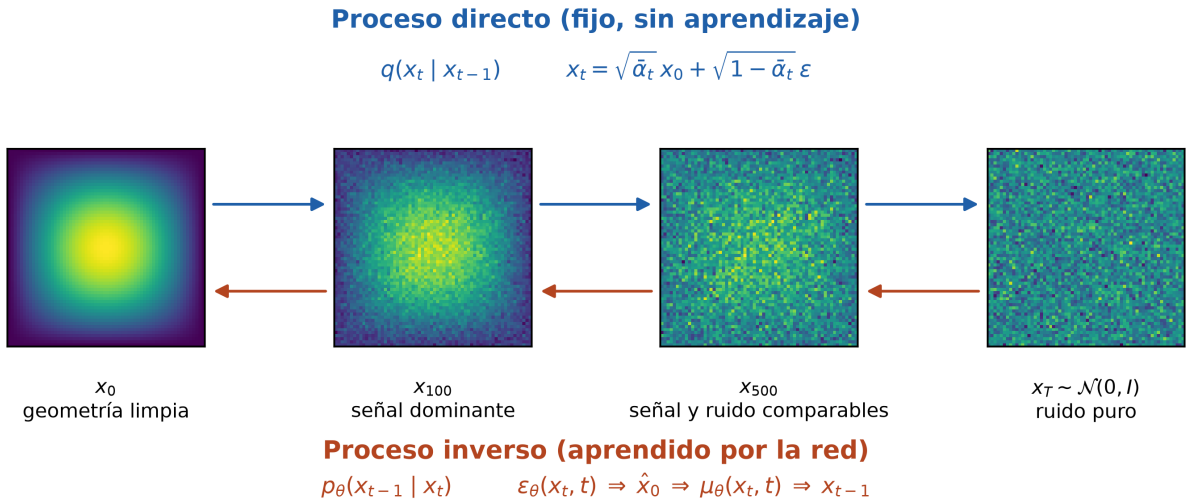

Figura 3.2: Esquema general de un modelo de difusión. El proceso directo transforma gradualmente una geometría limpia x_0 en ruido gaussiano x_T. El proceso inverso utiliza la predicción de la red para construir los estados $x_{T-1}, \ldots, x_0$.

Proceso directo de difusión

Sea x_0 una geometría limpia extraída de la distribución de datos. El proceso directo se define como una cadena de Markov

$$q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t \mid x_{t-1}), \tag{3.7}$$

en la que cada transición añade una pequeña cantidad de ruido gaussiano

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{\alpha_t}\, x_{t-1}, \beta_t I\right), \qquad \alpha_t = 1 - \beta_t. \tag{3.8}$$

El parámetro $\beta_t \in (0, 1)$ determina la cantidad de ruido introducida en el paso t. La multiplicación por $\sqrt{\alpha_t}$ reduce ligeramente la señal anterior, mientras que la varianza β_t incorpora una nueva perturbación. La misma transición puede escribirse mediante la reparametrización

$$x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{\beta_t}\, \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, I). \tag{3.9}$$

Aunque el proceso se define paso a paso, durante el entrenamiento no es necesario recorrer toda

la cadena para obtener x_t, ya que existe una expresión cerrada que genera directamente cualquier nivel de ruido a partir de x_0. Definiendo $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ y suponiendo por inducción que $x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\, x_0 + \sqrt{1-\bar{\alpha}_{t-1}}\, \varepsilon_{t-1}$, al sustituir en (3.9) se obtiene

$$x_t = \underbrace{\sqrt{\alpha_t \bar{\alpha}_{t-1}}}_{\sqrt{\bar{\alpha}_t}}\, x_0 + \sqrt{\alpha_t(1-\bar{\alpha}_{t-1})}\, \varepsilon_{t-1} + \sqrt{\beta_t}\, \varepsilon_t. \tag{3.10}$$

Los dos términos de ruido son gaussianos independientes, por lo que su suma también es gaussiana, con varianza total $\alpha_t(1-\bar{\alpha}_{t-1}) + \beta_t = \alpha_t - \alpha_t\bar{\alpha}_{t-1} + 1 - \alpha_t = 1 - \bar{\alpha}_t$. Por tanto, la distribución marginal de x_t condicionada a x_0 es

$$q(x_t \mid x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t) I\right), \tag{3.11}$$

y puede muestrearse directamente mediante

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I). \tag{3.12}$$

Esta expresión separa claramente las dos componentes del estado ruidoso. El término $\sqrt{\bar{\alpha}_t} x_0$ contiene la información que todavía se conserva de la geometría, mientras que $\sqrt{1-\bar{\alpha}_t}\varepsilon$ representa la perturbación añadida. Al comienzo del proceso, $\bar{\alpha}_t$ es próximo a uno y domina la señal. Al final, $\bar{\alpha}_t$ se aproxima a cero y el estado resulta prácticamente indistinguible de una muestra de $\mathcal{N}(0, I)$.

Cuando los datos se encuentran normalizados, la relación entre ambas contribuciones puede expresarse mediante

$$\mathrm{SNR}(t) = \frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}. \tag{3.13}$$

Una SNR alta indica que x_t todavía conserva gran parte de la geometría original. Una SNR baja indica que el ruido domina y que muchas geometrías distintas pueden producir estados compatibles con la observación.

Elección del schedule de ruido

La secuencia de valores $\beta_{t\,t=1}^{T}$ recibe el nombre de *noise schedule*. Su elección determina la velocidad con la que disminuye $\bar{\alpha}_t$ y, por tanto, cómo se distribuye la pérdida de información a lo largo de la trayectoria.

Un schedule lineal hace crecer β_t de forma uniforme entre dos valores prefijados. Aunque esta elección es sencilla, puede reducir la SNR con demasiada rapidez y dedicar numerosos pasos finales a estados donde apenas queda información de la geometría. Esos pasos aportan poco al aprendizaje del denoising y resultan especialmente poco útiles para aplicar un guiado mecánico.

En este trabajo se utiliza el schedule coseno propuesto por Nichol y Dhariwal (2021). En lugar de definir directamente una evolución lineal de β_t, se construye una trayectoria suave para $\bar{\alpha}_t$

$$f(t) = \cos^2\left(\frac{\frac{t}{T} + s}{1+s}\frac{\pi}{2}\right), \qquad \bar{\alpha}_t = \frac{f(t)}{f(0)}, \tag{3.14}$$

donde s es un pequeño desplazamiento que evita cambios excesivamente bruscos cerca del origen. A partir de $\bar{\alpha}_t$, las varianzas discretas se recuperan mediante

$$\beta_t = 1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}, \tag{3.15}$$

aplicando, cuando sea necesario, un límite superior para mantener la estabilidad numérica.

La trayectoria coseno mantiene valores intermedios de SNR durante una fracción mayor del proceso, como muestra la Figura 3.3. Esta propiedad es importante en el problema considerado, ya que la estructura global de la lámina sigue siendo reconocible durante más pasos. Como el guiado necesita simultáneamente suficiente información geométrica y suficiente libertad para modificar la muestra, esa región intermedia es la parte más útil de la trayectoria.

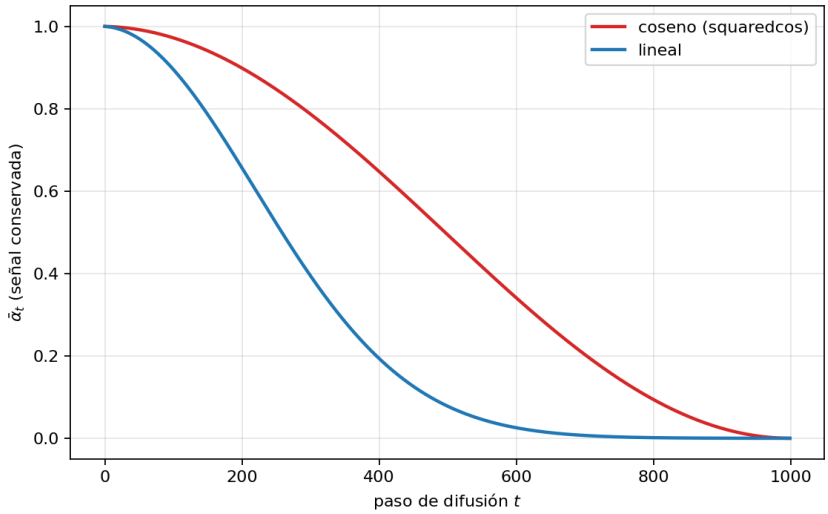

Figura 3.3: Evolución de la señal conservada $\bar{\alpha}_t$ a lo largo del proceso directo para los schedules lineal y coseno. El schedule coseno mantiene durante más tiempo estados intermedios que todavía contienen estructura geométrica aprovechable, que es la región donde el guiado resulta más efectivo.

3.2. Proceso inverso, entrenamiento y muestreo

Ambigüedad del problema inverso

La generación recorre la cadena en sentido contrario. Se parte de

$$x_T \sim \mathcal{N}(0, I), \tag{3.16}$$

y se construyen sucesivamente $x_{T-1}, x_{T-2}, \ldots, x_0$. El modelo generativo completo se expresa como

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t), \tag{3.17}$$

donde $p(x_T) = \mathcal{N}(0, I)$ y las transiciones inversas $p_\theta(x_{t-1} \mid x_t)$ se aproximan mediante una red neuronal.

La inversión no consiste simplemente en restar el ruido añadido. En el proceso directo se conoce tanto x_0 como la perturbación utilizada, pero durante la generación solo se observa x_t, y a medida que aumenta t el ruido elimina la información que distinguía unas geometrías de otras. En consecuencia, un mismo estado ruidoso puede ser compatible con varias geometrías limpias. La distribución $p(x_0 \mid x_t)$

describe todas las geometrías compatibles con la información presente en x_t. Para valores pequeños de t suele estar concentrada alrededor de una reconstrucción concreta, mientras que para valores altos puede ser ancha y presentar varios modos, correspondientes a familias geométricas diferentes que el ruido ya no permite distinguir, situación que ilustra la Figura 3.4.

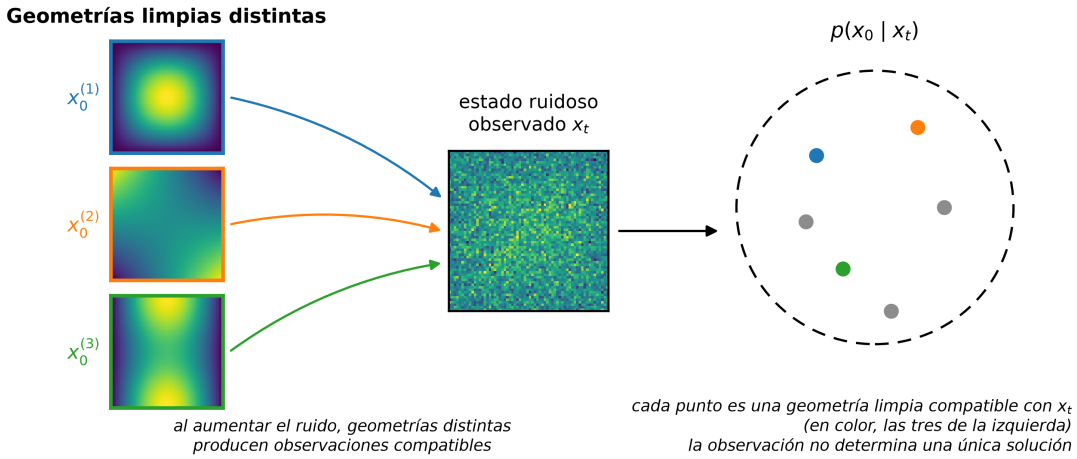

Figura 3.4: Ambigüedad del proceso inverso. Cuando el nivel de ruido es elevado, varias geometrías limpias pueden ser compatibles con la información contenida en un estado x_t. La distribución posterior $p(x_0 \mid x_t)$ no tiene por qué estar concentrada alrededor de una única solución.

Esta observación también corrige una interpretación frecuente de la convergencia al ruido. Dos geometrías corrompidas con perturbaciones independientes no terminan en la misma imagen de ruido. Producen realizaciones diferentes, pero esas realizaciones pertenecen aproximadamente a la misma distribución gaussiana y dejan de contener información suficiente para reconocer la geometría de origen. La Figura 3.5 muestra este comportamiento sobre láminas reales del conjunto de datos, siguiendo la corrupción progresiva de tres geometrías distintas.

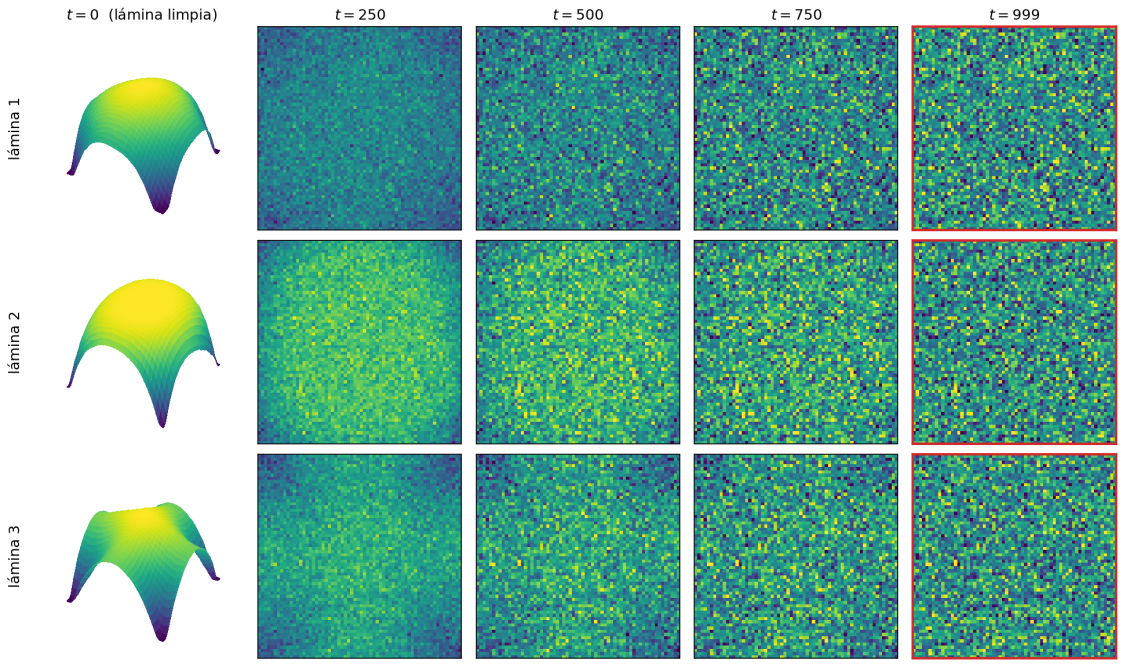

Figura 3.5: Pérdida progresiva de información durante el proceso directo. Tres láminas distintas ($t = 0$), corrompidas cada una con su propio ruido independiente, degeneran al aumentar t en realizaciones de ruido diferentes entre sí pero que ya no revelan qué geometría limpia originó cada una.

Posterior exacta condicionada a la geometría limpia

Aunque $p(x_0 \mid x_t)$ sea desconocida, el proceso directo permite calcular exactamente la transición

$$q(x_{t-1} \mid x_t, x_0). \tag{3.18}$$

Aplicando la regla de Bayes y eliminando los términos que no dependen de x_{t-1},

$$q(x_{t-1} \mid x_t, x_0) \propto q(x_t \mid x_{t-1}) q(x_{t-1} \mid x_0). \tag{3.19}$$

Las dos distribuciones del lado derecho son gaussianas conocidas. La primera viene dada por (3.8) y la segunda es la marginal (3.11) evaluada en $t-1$, es decir, $q(x_{t-1} \mid x_0) = \mathcal{N}(x_{t-1}; \sqrt{\bar{\alpha}_{t-1}}\, x_0, (1 - \bar{\alpha}_{t-1})I)$. El logaritmo negativo de su producto, ignorando las constantes independientes de x_{t-1}, es

$$\frac{1}{2}\left[\frac{|x_t - \sqrt{\alpha_t}x_{t-1}|^2}{\beta_t} + \frac{|x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}x_0|^2}{1-\bar{\alpha}_{t-1}}\right], \tag{3.20}$$

que es cuadrático en x_{t-1} y define por tanto otra gaussiana. Al desarrollar los cuadrados y agrupar, el término cuadrático da la precisión $1/\tilde{\beta}_t = \alpha_t/\beta_t + 1/(1-\bar{\alpha}_{t-1}) = (1-\bar{\alpha}_t)/\big(\beta_t(1-\bar{\alpha}_{t-1})\big)$, es decir,

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t, \tag{3.21}$$

y el término lineal da la media, que tras sustituir $\tilde{\beta}_t$ queda

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t. \tag{3.22}$$

En consecuencia,

$$q(x_{t-1} \mid x_t, x_0) = \mathcal{N}\left(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I\right). \tag{3.23}$$

Esta distribución es exacta, pero no puede utilizarse directamente durante la generación porque depende de x_0, que es precisamente la geometría desconocida. El papel de la red será proporcionar la información necesaria para aproximar esa dependencia.

Predicción del ruido

Existen diferentes parametrizaciones posibles para la red de difusión. Puede entrenarse para predecir directamente x_0, la media del paso inverso o el ruido añadido. En este trabajo se utiliza la parametrización mediante ruido introducida en los DDPM (Ho y cols., 2020).

A partir de una geometría limpia x_0, se selecciona un timestep t, se genera una perturbación

$$\varepsilon \sim \mathcal{N}(0, I), \tag{3.24}$$

y se construye x_t mediante (3.12). La red recibe (x_t, t) y produce una estimación $\varepsilon_\theta(x_t, t)$ del ruido

utilizado. La pérdida simplificada es

$$\mathcal{L}_{\text{diff}}(\theta) = \mathbb{E}_{x_0,t,\varepsilon}\left[|\varepsilon - \varepsilon_\theta(x_t,t)|^2\right]. \tag{3.25}$$

El significado del óptimo de esta pérdida se obtiene con la descomposición clásica del error cuadrático. Para cualquier predictor f,

$$\mathbb{E}\big[|\varepsilon - f(x_t,t)|^2\big] = \mathbb{E}\big[|\varepsilon - \mathbb{E}[\varepsilon \mid x_t,t]|^2\big] + \mathbb{E}\big[|f(x_t,t) - \mathbb{E}[\varepsilon \mid x_t,t]|^2\big], \tag{3.26}$$

donde el primer término no depende de f y el segundo es no negativo y solo se anula cuando f coincide con la esperanza condicional. Por tanto, en el límite de capacidad y datos suficientes,

$$\varepsilon_\theta(x_t,t) \longrightarrow \mathbb{E}[\varepsilon \mid x_t,t]. \tag{3.27}$$

La red no recupera necesariamente la perturbación concreta que produjo una muestra durante el proceso directo. Cuando x_t es ambiguo, predice el ruido medio compatible con ese estado.

Relación con el ELBO

La pérdida de ruido no es únicamente una regla empírica. El entrenamiento probabilístico del modelo puede formularse mediante una cota inferior de la log-verosimilitud, o ELBO,

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_q\Bigg[D_{\text{KL}}\left(q(x_T \mid x_0)\,\|\,p(x_T)\right) + \sum_{t=2}^{T} D_{\text{KL}}\left(q(x_{t-1} \mid x_t,x_0)\,\|\,p_\theta(x_{t-1} \mid x_t)\right) - \log p_\theta(x_0 \mid x_1)\Bigg].$$

El primer término no depende de la red y el último actúa como un término de reconstrucción del paso final. Los términos centrales son los que se entrenan. Si la transición aprendida se define como $p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1};\mu_\theta(x_t,t),\sigma_t^2 I)$, con la varianza fijada por el schedule, cada divergencia KL entre dos gaussianas de igual varianza se reduce, salvo constantes, al error cuadrático entre sus medias, $\|\tilde{\mu}_t - \mu_\theta\|^2/(2\sigma_t^2)$. Al parametrizar μ_θ mediante ε_θ, estos términos adoptan la forma

$w_t\,\mathbb{E}[|\varepsilon - \varepsilon_\theta(x_t, t)|^2]$, con un peso w_t que depende del timestep y del schedule. La pérdida de (3.25) elimina esa ponderación para equilibrar el aprendizaje entre niveles de ruido, por lo que puede entenderse como una versión simplificada del objetivo variacional original (Ho y cols., 2020).

Estimación de la geometría limpia

La ecuación directa puede despejarse para expresar la geometría limpia en función del estado y del ruido, $x_0 = \big(x_t - \sqrt{1-\bar{\alpha}_t}\,\varepsilon\big)/\sqrt{\bar{\alpha}_t}$. Tomando la esperanza condicionada a (x_t, t) y sustituyendo la esperanza del ruido, que es desconocida, por la predicción de la red, se obtiene

$$\hat{x}_0(x_t, t) = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\,\varepsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}} \approx \mathbb{E}[x_0 \mid x_t, t]. \tag{3.28}$$

Esta relación suele denominarse fórmula de Tweedie en el contexto de los modelos de difusión. La aproximación procede del error de la red. Si ε_θ alcanzase exactamente el óptimo de (3.27), $\hat{x}_0$ coincidiría con la media posterior.

Es importante interpretar correctamente esta cantidad. $\hat{x}_0$ no es necesariamente la única geometría limpia que originó x_t, sino el promedio de las geometrías compatibles con la observación. Cuando la posterior es estrecha, esta media puede representar una reconstrucción precisa. Cuando la posterior es ancha o multimodal, puede combinar características de varias soluciones y no corresponder exactamente a una geometría individual del conjunto de datos. Esta diferencia será relevante al estudiar el uso de surrogados mecánicos sobre estados ruidosos.

Construcción del paso inverso

La media exacta $\tilde{\mu}_t(x_t, x_0)$ de (3.22) depende de x_0. El modelo sustituye ese valor desconocido por la estimación $\hat{x}_0(x_t, t)$

$$\mu_\theta(x_t, t) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{x}_0(x_t, t) + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t. \tag{3.29}$$

Sustituyendo (3.28) y utilizando $\bar{\alpha}_t = \alpha_t\bar{\alpha}_{t-1}$, la misma media puede escribirse directamente en función de la predicción de ruido

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \varepsilon_\theta(x_t, t) \right). \tag{3.30}$$

El paso generativo se obtiene muestreando alrededor de esta media

$$x_{t-1} = \mu_\theta(x_t, t) + \sigma_t z, \qquad z \sim \mathcal{N}(0, I), \tag{3.31}$$

donde habitualmente se toma $\sigma_t^2 = \tilde{\beta}_t$ o una variante equivalente asociada al sampler. En el último paso se fija $z = 0$, evitando añadir ruido cuando se construye la muestra final. La Figura 3.6 resume la construcción completa del paso, desde la predicción de la red hasta el nuevo estado x_{t-1}.

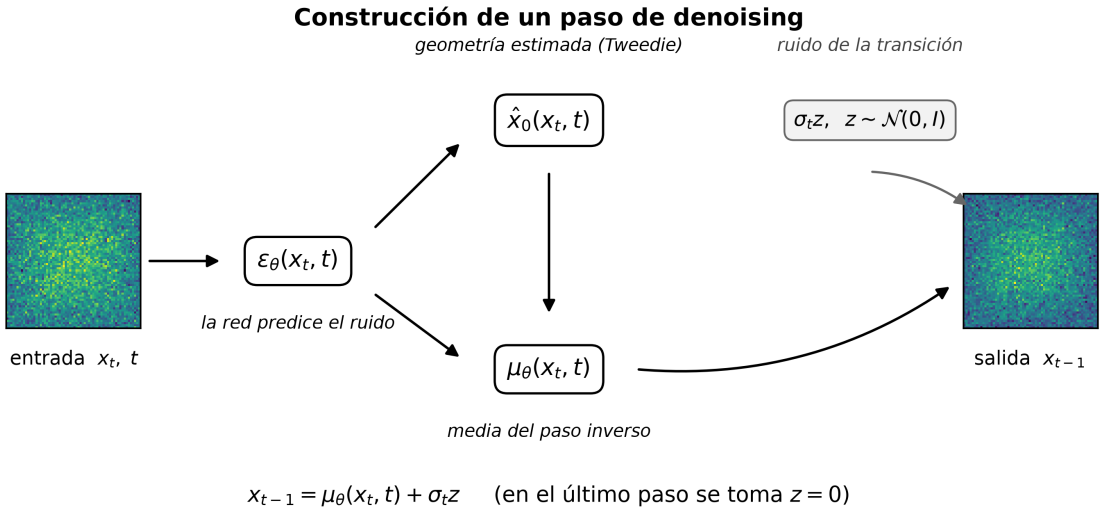

Figura 3.6: Construcción de un paso inverso. La red predice el ruido contenido en x_t. Esa predicción permite estimar la media posterior $\hat{x}_0$, calcular la media μ_θ de la transición inversa y obtener x_{t-1} añadiendo una perturbación de varianza controlada.

La componente aleatoria $\sigma_t z$ no constituye un error del procedimiento. Permite representar la incertidumbre de la transición y hace posible que distintas semillas produzcan geometrías diferentes. La red determina qué direcciones conducen hacia regiones plausibles de la distribución, mientras que el muestreo conserva la capacidad de explorar alternativas.

Relación entre predicción de ruido y score

La predicción de ruido también proporciona el *score* de la distribución ruidosa. Para una geometría limpia conocida, la distribución de x_t es la gaussiana de (3.11), cuyo score condicionado es $\nabla_{x_t} \log q(x_t \mid x_0) = -(x_t - \sqrt{\bar{\alpha}_t} x_0)/(1 - \bar{\alpha}_t)$. Como según (3.12) el numerador es exactamente $\sqrt{1 - \bar{\alpha}_t}\, \varepsilon$,

$$\nabla_{x_t} \log q(x_t \mid x_0) = -\frac{\varepsilon}{\sqrt{1 - \bar{\alpha}_t}}. \tag{3.32}$$

El score de la distribución marginal $p_t(x_t)$ se obtiene promediando este score condicionado sobre todas las geometrías compatibles con x_t

$$\nabla_{x_t} \log p_t(x_t) = \mathbb{E}\left[\nabla_{x_t} \log q(x_t \mid x_0) \mid x_t\right]. \tag{3.33}$$

Al combinar (3.32) con el óptimo de la predicción de ruido,

$$\nabla_{x_t} \log p_t(x_t) = -\frac{\mathbb{E}[\varepsilon \mid x_t, t]}{\sqrt{1 - \bar{\alpha}_t}} \approx -\frac{\varepsilon_\theta(x_t, t)}{\sqrt{1 - \bar{\alpha}_t}}. \tag{3.34}$$

Por tanto, entrenar la red para predecir ruido equivale a aprender el campo vectorial que indica, en cada nivel de ruido, hacia qué regiones debe desplazarse x_t para aumentar su probabilidad bajo la distribución de geometrías. Esta equivalencia conecta los DDPM con los modelos generativos basados en score y con su formulación continua mediante ecuaciones diferenciales estocásticas (Y. Song y cols., 2021), y es la que permite intervenir la trayectoria de muestreo con un gradiente físico, tal y como desarrolla la Sección 4.

3.3. Parametrización por velocidad

En lugar del ruido, el generador puede predecir la **velocidad** de difusión (Salimans y Ho, 2022),

$$v_t = \sqrt{\bar{\alpha}_t}\, \varepsilon - \sqrt{1 - \bar{\alpha}_t}\, x_0, \qquad \hat{x}_0 = \sqrt{\bar{\alpha}_t}\, x_t - \sqrt{1 - \bar{\alpha}_t}\, h_\theta(x_t, t), \tag{3.35}$$

con $h_\theta \approx v_t$. La razón de usarla es la estabilidad de la geometría implícita en el rango de ruido donde actúa el guiado, ya que usando la velocidad la varianza del error de $\hat{x}_0$ se amplifica por un factor acotado $(1 - \bar{\alpha}_t)$, mientras que bajo predicción de ruido se amplifica por $(1 - \bar{\alpha}_t)/\bar{\alpha}_t$, que diverge cuando $\bar{\alpha}_t \to 0$ (en norma del error, los factores son las raíces cuadradas de los anteriores, con la misma conclusión). Por eso la velocidad mantiene una geometría implícita mejor condicionada a ruido alto.

El proceso inverso no cambia por usar esta parametrización. En cada paso la red predice la velocidad $\hat{v} = h_\theta(x_t, t)$, y de ella se recuperan de forma exacta tanto la geometría limpia estimada como el ruido implícito, invirtiendo el sistema lineal (3.35),

$$\hat{x}_0 = \sqrt{\bar{\alpha}_t}\, x_t - \sqrt{1 - \bar{\alpha}_t}\, \hat{v}, \qquad \hat{\varepsilon} = \sqrt{1 - \bar{\alpha}_t}\, x_t + \sqrt{\bar{\alpha}_t}\, \hat{v}. \tag{3.36}$$

Con la geometría estimada $\hat{x}_0$ se ejecuta el mismo paso inverso (3.31), sustituyéndola en la media posterior $\tilde{\mu}_t$ y muestreando x_{t-1}. La parametrización por velocidad no altera, por tanto, la mecánica del proceso backward, solo la cantidad que la red regresa y, con ella, el condicionamiento numérico de $\hat{x}_0$ a ruido alto. Esta misma recuperación de $\hat{x}_0$ a partir de $\hat{v}$ es la que, en el guiado, permite propagar el gradiente físico hasta la velocidad (Sección 4).

4. Guiado físico como modificación del score

4.1. Condicionamiento mediante una energía

Hasta aquí el muestreo genera geometría plausible, es decir, muestrea de la distribución que ha aprendido el prior. Pero el objetivo de este trabajo no es solo generar láminas verosímiles, sino láminas que además cumplan una propiedad mecánica, que se denotará por Φ y que en la práctica será tener un factor de membrana alto. Dicho de otro modo, se quiere muestrear de la distribución **condicionada** $p(x_t \mid \Phi)$ en lugar de la distribución sin condicionar $p_t(x_t)$.

El puente entre ambas lo da el teorema de Bayes, que reescribe la probabilidad condicionada en función de cantidades que sí se saben modelar,

$$p(x_t \mid \Phi) = \frac{p(\Phi \mid x_t)\, p_t(x_t)}{p(\Phi)}. \tag{3.37}$$

El primer factor del numerador, $p(\Phi \mid x_t)$, mide cómo de bien cumple un estado x_t la propiedad deseada. El segundo, $p_t(x_t)$, es la distribución que ya conoce el prior. El denominador $p(\Phi)$ es una constante de normalización que no depende de x_t.

El muestreo por difusión no necesita la densidad en sí, sino su **score**, que es el gradiente del logaritmo de la densidad respecto del estado, $\nabla_{x_t} \log p$. El score apunta en la dirección en la que la densidad crece, esto es, hacia los estados más probables, y es la cantidad que la red de difusión aprende a estimar (Sección 3). Tomando logaritmos en la ecuación (3.37) y derivando respecto de x_t, el producto se convierte en suma y la constante $p(\Phi)$ se anula por tener gradiente nulo, de modo que

$$\nabla_{x_t} \log p(x_t \mid \Phi) = \underbrace{\nabla_{x_t} \log p_t(x_t)}_{\text{score del prior}} + \underbrace{\nabla_{x_t} \log p(\Phi \mid x_t, t)}_{\text{corrección física}}. \tag{3.38}$$

Esta descomposición (Dhariwal y Nichol, 2021; Ho y Salimans, 2022) es la idea central del guiado. Muestrear geometría condicionada equivale a seguir el score del prior, que ya se tiene, más un término de corrección que empuja hacia los estados que cumplen la propiedad. El primer sumando lo aporta el Arquitecto. El segundo es el que hay que construir a partir de la física, y es donde entra el Ingeniero.

Queda, por tanto, dar forma a $p(\Phi \mid x_t, t)$, la probabilidad de que un estado cumpla la propiedad. En este trabajo esa propiedad se expresa como un **coste** $J(x_t)$ que vale poco cuando el estado es bueno, con factor de membrana alto, y mucho cuando es malo. Convertir un coste en una probabilidad es un problema clásico que resuelve la distribución de Boltzmann, también llamada de Gibbs, que asigna a cada estado una probabilidad que decae de forma exponencial con su energía,

$$p(\Phi \mid x_t, t) = \frac{1}{Z}\, e^{-\lambda J(x_t)}, \qquad Z = \int e^{-\lambda J(x)}\, dx. \tag{3.39}$$

La lectura es intuitiva. Los estados de coste bajo reciben probabilidad alta y los de coste alto, probabilidad baja. El parámetro $\lambda > 0$ hace de inverso de la temperatura y regula cómo de tajante es esa preferencia, ya que un valor grande concentra casi toda la probabilidad en los estados de coste mínimo y un valor pequeño la reparte de forma más uniforme. Esta forma exponencial no es arbitraria, sino la distribución de máxima entropía compatible con un valor esperado dado de la

energía (Jaynes, 1957), es decir, la elección menos comprometida que respeta lo único que se impone sobre el sistema. El término Z, llamado función de partición, normaliza la distribución y tampoco depende de x_t. Se escribe λ en lugar del habitual β para no confundirlo con el schedule de varianzas β_t de la difusión.

Con esto ya se puede calcular la corrección física de forma explícita. Tomando el logaritmo de la ecuación (3.39) se obtiene

$$\log p(\Phi \mid x_t, t) = -\lambda J(x_t) - \log Z, \tag{3.40}$$

y al derivar respecto de x_t el término $\log Z$ desaparece por ser constante en x_t, con lo que queda

$$\nabla_{x_t} \log p(\Phi \mid x_t, t) = -\lambda \nabla_{x_t} J(x_t). \tag{3.41}$$

El resultado tiene un significado muy concreto. La corrección física es, salvo la constante λ, el gradiente del coste con el signo cambiado, es decir, la dirección de **descenso** del coste. Sustituida en la ecuación (3.38), empuja en cada paso la trayectoria de muestreo hacia estados de menor coste, esto es, de mayor factor de membrana, sin dejar de seguir el score del prior que mantiene la geometría plausible. En la práctica la constante λ no se ajusta de forma independiente, sino que se absorbe en un único escalar, la fuerza de guiado, que se estudia en los resultados.

En las láminas, el coste J se calcula con el factor de membrana que predice el surrogado, y su gradiente $\nabla_{x_t} J$ se obtiene por *backpropagation* a través del Ingeniero, por ser una red diferenciable. Así, la física entra en la generación como un término aditivo sobre el score, calculado por el mismo surrogado que evalúa la mecánica.

4.2. Acoplamiento con la trayectoria

El gradiente físico define una dirección de descenso $-\nabla_{x_t} J$ sobre la geometría. Al acoplarlo a la predicción de velocidad h_θ hay que cuidar el signo, porque la velocidad entra en la geometría implícita con signo negativo, como se ve en (3.35), donde $\hat{x}_0 = \sqrt{\bar{\alpha}_t}\, x_t - \sqrt{1-\bar{\alpha}_t}\, h_\theta$, de modo que **aumentar** la velocidad a lo largo de $+\nabla_{x_t} J$ desplaza la geometría implícita a lo largo de $-\nabla_{x_t} J$, que es lo que se busca. La corrección se expresa por tanto en el espacio de velocidad como $\tilde{g}_t \propto +\nabla_{x_t} J$, escalada,

recortada y ponderada en el tiempo, y de la conversión entre espacios aparece el factor $\sqrt{1-\bar{\alpha}_t}$,

$$h_{\text{guided}}(x_t, t) = h_\theta(x_t, t) + \sqrt{1-\bar{\alpha}_t}\,\tilde{g}_t. \tag{3.42}$$

El esquema base concentra la corrección con una envolvente gaussiana sobre el paso de muestreo,

$$w_k = w_{\text{máx}} \exp\left(-\frac{(r_k-\rho)^2}{2\,\sigma_{\text{bell}}^2}\right), \qquad r_k = \frac{k}{K-1}, \tag{3.43}$$

con anchura σ_{bell}, centro ρ y máximo $w_{\text{máx}}$. La intuición es guiar en ruido intermedio con la idea de que al inicio la muestra no tiene estructura interpretable y al final la geometría ya está decidida. El resto del capítulo motiva por qué esa forma fija puede sustituirse por una derivada de la incertidumbre del surrogado.

5. Jensen, media condicional y frontera de información

5.1. Desajuste de Jensen a nivel de operador

El muestreo visita estados ruidosos x_t, pero el operador mecánico está definido sobre geometría limpia. Aquí aparece el hecho que recorre todo el capítulo, y es que **un mismo x_t es compatible con muchas geometrías limpias x_0 distintas**, sobre todo a ruido medio y alto, porque el estado observado no las distingue. La estrategia habitual en difusión es estimar una geometría limpia por Tweedie (3.28) y evaluar el operador sobre ella (Chung y cols., 2023; Efron, 2011), es decir $R(\hat{x}_0, f_z)$. Pero R es no lineal, así que

$$R(\mathbb{E}[x_0 \mid x_t, t], f_z) \neq \mathbb{E}[R(x_0, f_z) \mid x_t, t]\,. \tag{3.44}$$

Esta diferencia es el **desajuste de Jensen a nivel de operador**, es decir, evaluar la mecánica sobre la media de las geometrías plausibles no equivale a la media de sus respuestas.

5.2. El surrogado como media condicional

La forma de evitarlo es entrenar el surrogado directamente sobre los estados ruidosos que verá el muestreo. Con pérdida cuadrática contra la respuesta FEM limpia,

$$\mathcal{L}(\phi) = \mathbb{E}\big[\|R(x_0, f_z) - g_\phi(x_t, f_z, t)\|^2\big]\,, \tag{3.45}$$

el minimizador ideal (datos, capacidad y optimización suficientes) es la esperanza condicional

$$g^\star_\phi(x_t, f_z, t) = \mathbb{E}[R(x_0, f_z) \mid x_t, f_z, t]\,. \tag{3.46}$$

La consecuencia, que es la que desarrolla esta memoria, es que si el surrogado ideal ya devuelve la media condicional, el error que le queda no es necesariamente falta de capacidad, sino información que la difusión ha destruido.

5.3. Frontera de información y varianza posterior

Para un x_t fijo, el error cuadrático esperado de cualquier surrogado se descompone como

$$\mathbb{E}\big[\|R - g_\phi\|^2 \mid x_t, t\big] = \underbrace{\big\|\mathbb{E}[R \mid x_t, t] - g_\phi\big\|^2}_{\text{sesgo}^2} + \mathrm{Tr}\big(\Sigma_R(x_t, t)\big), \qquad \Sigma_R = \mathrm{Var}[R \mid x_t, t]. \tag{3.47}$$

La descomposición es exacta porque g_ϕ es función determinista de x_t, de modo que el término cruzado se anula. En el límite ideal $g_\phi = g^\star_\phi$ el sesgo desaparece y queda la traza de la **varianza posterior** Σ_R. Esta es la **frontera de información** que ningún surrogado puede recuperar de x_t lo que el ruido ha ocultado. La consecuencia metodológica es que si en una ventana de timesteps el error está dominado por Σ_R y no por el sesgo, añadir parámetros o un sesgo arquitectónico más complejo rinde poco, de modo que la pregunta correcta es qué arquitectura aproxima la media condicional con el coste adecuado y qué incertidumbre da sobre la parte no recuperable.

5.4. Brecha residual a nivel del objetivo

La media condicional mitiga la brecha de Jensen en el operador, pero el objetivo de guiado sigue siendo no lineal (el factor de membrana es un cociente). Por tanto $\mathbb{E}[J(R) \mid x_t, t] \neq J(\mathbb{E}[R \mid x_t, t])$.

Con $\delta = R - g_\phi$ y un desarrollo de Taylor de segundo orden,

$$\mathbb{E}[J(g_\phi + \delta) - J(g_\phi) \mid x_t, t] \approx \underbrace{\nabla J(g_\phi)^\top \mathbb{E}[\delta \mid x_t, t]}_{T_1} + \tfrac{1}{2} \underbrace{\mathbb{E}[\delta^\top \nabla^2 J(g_\phi)\, \delta \mid x_t, t]}_{T_2}. \tag{3.48}$$

Si el surrogado alcanza la media condicional, $g_\phi = g_\phi^\star$, entonces $\mathbb{E}[\delta \mid x_t, t] = 0$, el término lineal T_1 se anula, y el residuo queda

$$T_2 \approx \tfrac{1}{2} \operatorname{Tr}\big(\nabla^2 J(g_\phi^\star)\, \Sigma_R(x_t, t)\big). \tag{3.49}$$

Es decir, la brecha de Jensen del objetivo es la interacción entre la curvatura de J y la misma varianza posterior Σ_R que aparece en (3.47). La aportación de esta memoria es precisamente esa visión unificada, en la que el sesgo del guiado y el error irreducible del surrogado quedan gobernados por el mismo objeto. La Figura 3.7 esquematiza el desajuste, comparando la ruta que promedia las geometrías antes de aplicar el operador con la que aplica el operador a cada geometría y promedia después las respuestas.

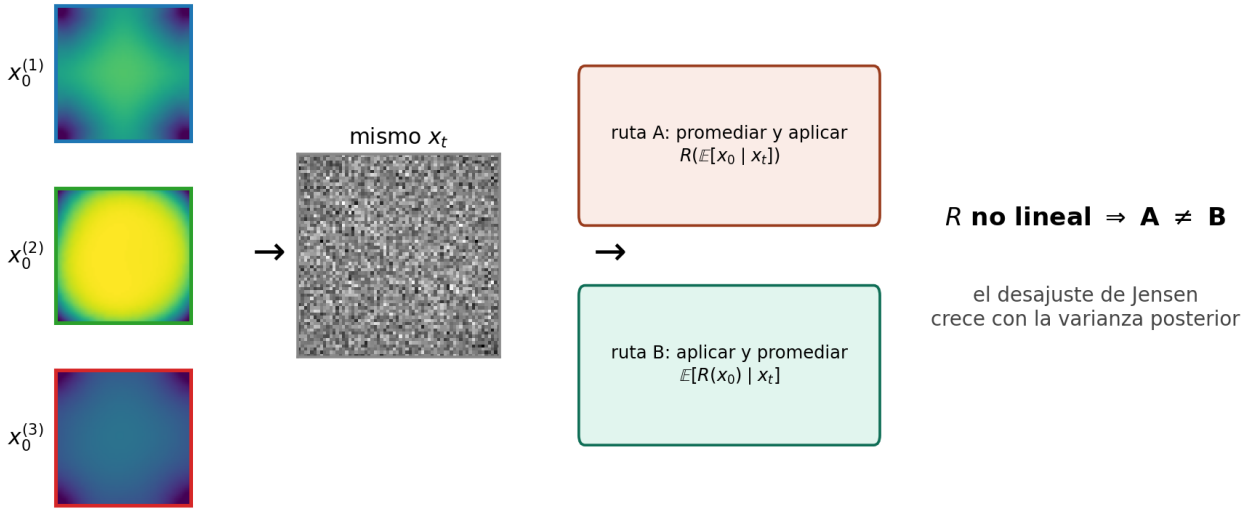

Figura 3.7: Esquema del desajuste de Jensen. Varias geometrías limpias distintas son compatibles con un mismo estado ruidoso x_t. Como el operador mecánico R es no lineal, promediar las geometrías y aplicar R (ruta A) no equivale a aplicar R a cada geometría y promediar las respuestas (ruta B). El surrogado consciente del ruido aprende directamente la ruta B; el residuo restante (3.49) es curvatura por varianza posterior.

6. Análisis espectral mediante SNR y filtro de Wiener

La sección anterior deja Σ_R como objeto central, pero abstracto. La transformada de Fourier le da forma cerrada y, de paso, explica qué arquitectura de surrogado conviene.

6.1. SNR por modo y varianza posterior en forma cerrada

La sección anterior dejó como pieza central la varianza posterior $\mathrm{Var}[x_0 \mid x_t]$, que mide cuánta incertidumbre permanece sobre la geometría limpia x_0 una vez observado el estado ruidoso x_t. En el espacio de píxeles esta varianza es una matriz de gran tamaño y con sus componentes acopladas, poco manejable analíticamente. La observación que organiza toda la sección es que, al llevar el problema al dominio de Fourier, esa matriz **se diagonaliza**, de modo que cada modo, es decir, cada frecuencia espacial, se desacopla de los demás y admite una expresión escalar propia. Esto es lo que se entiende por forma cerrada, una fórmula sencilla modo a modo en lugar de una matriz acoplada.

El punto de partida es la relación directa de la difusión en el espacio real, la ecuación (3.12), en la que el estado ruidoso es la geometría limpia atenuada más ruido gaussiano, $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon$. Como la transformada de Fourier es lineal, actúa sobre cada término por separado y cada frecuencia k evoluciona de forma independiente,

$$X_t(k) = \sqrt{\bar{\alpha}_t}\, X_0(k) + \sqrt{1-\bar{\alpha}_t}\, E(k), \tag{3.50}$$

donde $X_t(k)$, $X_0(k)$ y $E(k)$ son los coeficientes de Fourier de x_t, x_0 y ε a la frecuencia k. La relación es ya escalar y los distintos modos no interactúan entre sí, y en ese desacoplamiento reside la ventaja de trabajar en el dominio de la frecuencia.

El análisis se apoya en dos hipótesis. La primera es que el ruido es blanco, $\mathbb{E}|E(k)|^2 = 1$ para todo k, de manera que reparte la misma potencia en todas las frecuencias y presenta un espectro plano, tal como corresponde a la transformada de un ruido gaussiano independiente en cada píxel. La segunda es la definición del espectro de potencia de las láminas, $S(k) = \mathbb{E}|X_0(k)|^2$, que expresa cuánta energía de señal contiene la geometría limpia en la frecuencia k. Los campos físicos reales se comportan casi siempre como un *low-pass filter*, con mucha energía en las frecuencias bajas, que recogen la forma global y suave, y poca en las altas, asociadas al detalle fino, y este hecho es el que determina la

conclusión de la sección. Se supone además, como es habitual, que señal y ruido están incorrelados.

Con estos ingredientes, la relación señal-ruido de cada modo se define como el cociente entre la potencia de la componente de señal y la de la componente de ruido dentro de la observación $X_t(k)$. De la ecuación (3.50), la señal es $\sqrt{\bar{\alpha}_t}\, X_0(k)$, con potencia $\bar{\alpha}_t\, S(k)$, y el ruido es $\sqrt{1-\bar{\alpha}_t}\, E(k)$, con potencia $1-\bar{\alpha}_t$, de modo que su cociente es

$$\mathrm{SNR}_t(k) = \frac{\bar{\alpha}_t\, S(k)}{1-\bar{\alpha}_t}. \tag{3.51}$$

La expresión se interpreta de forma directa. Una mayor energía de señal en el modo o una menor cantidad de ruido inyectado elevan la relación señal-ruido, mientras que un ruido mayor la reduce. A medida que t crece y el estado se vuelve más ruidoso, $\bar{\alpha}_t$ tiende a cero y la relación señal-ruido de todos los modos decae hacia cero, de forma que en el límite la señal queda por completo dominada por el ruido.

Queda estimar la geometría limpia a partir de la observación. Recuperar $X_0(k)$ dado $X_t(k)$ es un problema de estimación gaussiana escalar, que tiene solución en forma cerrada. Con un prior $X_0(k) \sim \mathcal{N}(0, S(k))$ y una observación $X_t(k) = \sqrt{\bar{\alpha}_t}\, X_0(k) + \text{ruido de varianza } 1-\bar{\alpha}_t$, la varianza posterior, con varianza de prior $S(k)$, ganancia de medida $a = \sqrt{\bar{\alpha}_t}$ y varianza de ruido $\sigma^2 = 1-\bar{\alpha}_t$, es

$$\mathrm{Var}[X_0(k) \mid X_t(k)] = \left(\frac{1}{S(k)} + \frac{a^2}{\sigma^2}\right)^{-1} = \left(\frac{1}{S(k)} + \frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}\right)^{-1}, \tag{3.52}$$

y, poniendo a común denominador y tomando el inverso, se obtiene la forma cerrada

$$\mathrm{Var}[X_0(k) \mid X_t(k)] = \frac{(1-\bar{\alpha}_t)\, S(k)}{\bar{\alpha}_t\, S(k) + 1 - \bar{\alpha}_t}. \tag{3.53}$$

Tanto la relación señal-ruido como la varianza posterior se siguen por derivación de la estimación gaussiana escalar, y no se introducen como postulados.

La varianza posterior admite dos lecturas según el régimen de ruido. Cuando $\mathrm{SNR}_t(k) \gg 1$, el modo conserva mucha señal frente al ruido, el término $\bar{\alpha}_t S(k)$ domina el denominador y la varianza tiende a $(1-\bar{\alpha}_t)/\bar{\alpha}_t$, que se anula, de manera que el modo se recupera sin incertidumbre residual. Cuando $\mathrm{SNR}_t(k) \ll 1$, el modo queda dominado por el ruido, el término $1-\bar{\alpha}_t$ domina el denominador

y la varianza vuelve a valer $S(k)$, la del prior, lo que indica que la observación no ha aportado información y que el mejor estimador es la media del prior, nula en este caso. El modo se pierde.

Al sumar la varianza posterior sobre todos los modos se recupera la traza de la matriz de covarianza posterior, $\mathrm{Tr}(\mathrm{Var}[x_0 \mid x_t])$, un único número que resume la incertidumbre total que permanece sobre la geometría limpia al nivel de ruido t. Es la **frontera de información** introducida antes, la porción de x_0 que sigue siendo recuperable a partir de x_t. El paso al dominio de Fourier es lo que convierte esa traza, que en principio exigiría los autovalores de una matriz grande, en una simple suma escalar sobre los modos.

De todo lo anterior se desprende la idea que vertebra la sección, **los modos altos caen antes por debajo del ruido, son irrecuperables, y el estimador óptimo los atenúa en lugar de reconstruirlos**. El argumento encadena los pasos previos. Como el espectro $S(k)$ de un campo físico decae con la frecuencia, los modos altos parten de poca energía de señal y, por la fórmula de la relación señal-ruido, cruzan el umbral del ruido antes que los bajos, es decir, a niveles de ruido menores. Por el análisis de los dos límites, esos modos de relación señal-ruido baja tienen una varianza posterior próxima a $S(k)$ y son, por tanto, irrecuperables. El estimador óptimo, el filtro de Wiener, no intenta reconstruirlos, sino que pondera cada modo por el factor

$$\frac{\mathrm{SNR}_t(k)}{1+\mathrm{SNR}_t(k)}, \tag{3.54}$$

próximo a uno cuando la relación señal-ruido es alta, con lo que deja pasar el modo, y próximo a cero cuando es baja, con lo que lo atenúa. La consecuencia de diseño, que es el objetivo último de la sección, es que un operador espectral, que trabaja modo a modo y concentra su capacidad en los modos informativos, puede igualar a una red convolucional que procesa todo el campo por igual empleando muchos menos parámetros, porque el propio análisis anticipa que los modos altos son ruido irrecuperable y que dedicar capacidad a reconstruirlos es un desperdicio. La arquitectura espectral incorpora ese conocimiento sin coste adicional.

6.2. De la geometría a la respuesta, y los límites del argumento

El objetivo no depende de x_0 sino de $R(x_0, f_z)$. Linealizando el operador, $\Sigma_R \approx J_R \,\mathrm{Var}[x_0 \mid x_t, t]\, J_R^\top$, con J_R el Jacobiano de R. La mecánica cierra el círculo, ya que **la flexión depende de curvaturas, que amplifican los modos altos**, justo los de menor SNR y mayor varianza residual en las ecuaciones (3.51) y (3.53). La parte flexional es por tanto la más difícil de recuperar de un estado ruidoso, y el factor de membrana (un cociente entre membrana y flexión) arrastra una componente de error irreducible.

Este resultado no demuestra que el operador espectal sea la mejor arquitectura posible, sino que puede ser más eficiente cuando la magnitud que se quiera predecir depende de patrones globales y de baja frecuencia. En cambio, si lo que se pretende es reconstruir con precisión los campos mecánicos completos, incluyendo bordes, discontinuidades y concentraciones locales de tensión, una red convolucional profunda puede ser más adecuada. Es por eso entendible que el operador espectral reduzca el error en MF-MAE, pero presente mayor error al reconstruir los campos físicos, medido mediante PhysMSE. La ventaja aparece cuando la información de interés está concentrada en unos pocos modos de baja frecuencia que siguen siendo identificables incluso con ruido.

Esta pérdida progresiva de información espectral se analizará empíricamente en el capítulo de resultados, en la Figura 5.4. En ella se observa que el último modo recuperable disminuye desde aproximadamente $k^\star \approx 22$ en el estado limpio hasta $k^\star \approx 1$ en la región de máximo ruido utilizada durante el guiado.

7. Incertidumbre del surrogado y su calibración

Hasta aquí Σ_R ha sido un objeto teórico, la varianza de las respuestas mecánicas compatibles con un estado ruidoso, que marca la parte del error que ningún surrogado puede eliminar. La pregunta práctica es cómo conocer ese límite sin tener acceso a él, y la respuesta de este trabajo es hacer que la propia red lo estime. La idea consiste en que el surrogado no devuelva solo su mejor predicción μ_ϕ, sino también un segundo número por píxel, σ_ϕ^2, que representa cuánto espera equivocarse en ese punto. Una varianza que depende de la entrada se denomina **heterocedástica**, en contraste con la suposición habitual de un error del mismo tamaño en todas partes, y es exactamente lo que se

necesita aquí, ya que el error real cambia con el nivel de ruido del estado x_t.

Para entrenar la media y la varianza a la vez hace falta una pérdida que las acople, y esa pérdida no se postula, sino que se deriva de un supuesto probabilístico sencillo. Se modela la respuesta verdadera y como una gaussiana centrada en la predicción de la red y con la varianza que la propia red declara,

$$p(y \mid x_t, f_z, t) = \frac{1}{\sqrt{2\pi\,\sigma_\phi^2}} \exp\left(-\frac{(y-\mu_\phi)^2}{2\,\sigma_\phi^2}\right). \tag{3.55}$$

Entrenar por máxima verosimilitud significa elegir los pesos que hacen más probables los datos observados, lo que equivale a minimizar el logaritmo negativo de esa densidad. Tomando $-\log$ en (3.55) y descartando el término constante $\frac{1}{2}\log 2\pi$, que no depende de la red, queda

$$-\log p(y \mid x_t, f_z, t) = \frac{(y-\mu_\phi)^2}{2\,\sigma_\phi^2} + \tfrac{1}{2}\log\sigma_\phi^2 + \text{const.} \tag{3.56}$$

Para garantizar que la varianza sea positiva sin imponer restricciones, la red predice en realidad su logaritmo $\ell_\phi = \log\sigma_\phi^2$, y sustituyendo en (3.56) se obtiene la pérdida empleada,

$$\mathcal{L}_{\text{NLL}} = \tfrac{1}{2}\big(\exp(-\ell_\phi)(y-\mu_\phi)^2 + \ell_\phi\big), \tag{3.57}$$

aplicada por píxel y canal y promediada sobre el lote.

Esta pérdida se autorregula gracias al equilibrio entre sus dos términos. El primero divide el error cuadrático por la varianza declarada, de modo que anunciar mucha incertidumbre amortigua el castigo de un error grande. El segundo penaliza precisamente esa declaración, impidiendo que la red se escude en una incertidumbre enorme para todo. El compromiso óptimo se encuentra derivando la pérdida esperada respecto de ℓ_ϕ e igualando a cero, lo que da

$$\sigma_\phi^{2\star} = \mathbb{E}\big[(y-\mu_\phi)^2 \mid x_t, t\big], \tag{3.58}$$

es decir, la varianza que aprende la red converge al error cuadrático que su propia media comete de verdad en ese punto. Y aquí se cierra el hilo del capítulo, porque cuando la media alcanza la respuesta condicional de la Sección 5, ese error cuadrático es justamente la diagonal de la misma Σ_R

que limita el error (3.47) y gobierna el residuo de Jensen (3.49), de modo que la σ^2 que aprende la red no es un añadido decorativo, sino una estimación operativa de la frontera de información.

Conviene precisar qué tipo de incertidumbre es esta. La que estima σ_ϕ^2 procede del propio ruido del estado, ya que muchas geometrías limpias distintas son compatibles con el mismo x_t (la multiplicidad de la Sección 5) y ni el mejor modelo posible podría distinguirlas, por lo que no disminuye entrenando con más datos. En la literatura se la llama incertidumbre **aleatoria**, frente a la **epistémica**, que refleja lo que el modelo ignora por haber visto pocos ejemplos y sí se reduce con más datos o promediando varias redes entrenadas de forma independiente (Gal y Ghahramani, 2016; Kendall y Gal, 2017; Lakshminarayanan, Pritzel, y Blundell, 2017).

Queda una última exigencia, y es que predecir una varianza no garantiza que sea creíble. Se dice que la incertidumbre está **calibrada** cuando los números que declara la red se corresponden con los errores que de verdad se observan, es decir, cuando en los puntos donde predice una varianza de valor v el error cuadrático medio ronda efectivamente v. La comprobación práctica consiste en agrupar las predicciones según su varianza declarada, medir el error real dentro de cada grupo y dibujar ambas cantidades una frente a otra, quedando un modelo bien calibrado sobre la diagonal (Kuleshov, Fenner, y Ermon, 2018; Levi y cols., 2019). El caso con entrada ruidosa es especialmente delicado, porque el ruido se propaga de forma no lineal por el operador mecánico (Zou, Meng, y Karniadakis, 2023).

En esta memoria la calibración se estudia a lo largo del tiempo de difusión y se le piden dos cosas distintas. La primera es de orden, ya que la σ^2 debe crecer y decrecer al compás del error real conforme cambia el nivel de ruido. La segunda es de escala, porque una red puede ordenar bien el error y aun así quedarse corta o pasarse en su magnitud. Lo primero no puede arreglarse a posteriori, pero lo segundo sí, multiplicando todas las varianzas por un único factor τ ajustado en validación, $\tilde{\sigma}^2 = \tau\,\sigma_\phi^2$, una técnica conocida como escalado por temperatura. Una corrección más fina, con un factor y un desplazamiento propios de cada timestep, $\tilde{\sigma}_t^2 = a_t\sigma_t^2 + b_t$, es el refinamiento natural y se deja como extensión.

8. De la incertidumbre al guiado adaptativo

La campana (3.43) guía en ruido intermedio, pero su pico, anchura y centro son hiperparámetros que no dependen de la fiabilidad real del surrogado en cada paso. La pregunta central de esta memoria es si esa fiabilidad (ya estimada por σ_ϕ^2) puede fijar la forma del guiado. La respuesta principal es temporal, calculando la incertidumbre por paso, $u_t = \mathbb{E}_{\Omega,c}[\sigma_\phi^2(x_t, t)]$, y se define un peso inversamente proporcional a ella, normalizado para igualar el presupuesto total de la campana,

$$w_t^\sigma = C\,\frac{1}{u_t + \epsilon}, \qquad \sum_t w_t^\sigma = \sum_t w_t^{\text{bell}}. \tag{3.59}$$

Así, la incertidumbre no decide **qué** píxel corregir, sino **cuándo** merece escucharse el surrogado, pues el valor de σ^2 está en el "cuándo" del guiado.

Existe una tensión que conviene explicitar. La fiabilidad crece al bajar el ruido, pero la maleabilidad de la geometría disminuye porque la trayectoria ya está cerca de una muestra limpia. Un schedule basado solo en $1/u_t$ prioriza la fiabilidad y puede actuar algo tarde para reformar la geometría, y por eso cabe esperar que se acerque mucho a una campana bien afinada sin igualarla en todos los regímenes, cambiando ajuste manual por interpretabilidad. Conviene notar que el mapa espacial completo $\sigma_\phi^2(x, t)$ abriría además la pregunta distinta de **dónde** corregir dentro de cada geometría. Este trabajo se centra en el **cuándo**, que es donde la lectura espectral sitúa la información dominante (a qué nivel de ruido el gradiente es informativo, no cómo se reparte espacialmente), y deja la explotación del mapa espacial como línea de trabajo futuro.

La forma concreta que adopta este schedule frente a la campana, deducida del perfil de incertidumbre medido, se muestra en la Figura 4.6 del capítulo de metodología.

Capítulo 4

Metodología

Este capítulo describe cómo se construyen, entrenan y evalúan los elementos que sostienen la propuesta. El hilo del capítulo es sencillo. Primero se fija la representación de los datos. Después se presenta el pipeline de dos redes, un prior geométrico de difusión y un surrogado mecánico consciente del ruido. A continuación se detallan las arquitecturas de surrogado comparadas y su entrenamiento. Y por último se explica cómo la incertidumbre del surrogado se convierte en una regla de guiado y cómo se evalúa todo el sistema. El prior geométrico se mantiene congelado a lo largo del trabajo, de modo que cualquier diferencia medida se atribuye al surrogado o a la estrategia de guiado, y no al generador de formas.

Para mantener la notación alineada con el capítulo anterior, se usa w para los pesos temporales de guiado, σ_{bell} para la anchura de la campana y $\omega_{\text{phys}}(t)$ para el peso temporal de la regularización física, evitando sobrecargar un mismo símbolo con dos significados.

1. Datos: representación geométrica y mecánica

El conjunto de datos procede del trabajo previo de Lourenço et al. (Lourenço y cols., 2024) y está formado por simulaciones de láminas funiculares discretizadas en una rejilla fija de 64×64. Las geometrías se generan replicando computacionalmente la analogía del modelo colgante. Un plano cuadrado de 10×10 m^2, fijado en sus esquinas y con o sin hueco central, se somete a una simulación de tipo *cloth* sin rigidez a flexión bajo la acción de la gravedad, y la superficie invertida resultante

es la candidata a forma funicular. La variabilidad se introduce mediante puntos de control sobre el dominio. La respuesta mecánica de referencia se obtiene después con un análisis estático lineal de lámina delgada en Abaqus/Standard, con elementos de lámina S4R y S3R de tamaño aproximado 0,25 m y condiciones de contorno fijas en los nodos de las esquinas en contacto con el suelo. El material es un hormigón elástico lineal isótropo cuyas propiedades, junto con las del problema, se resumen en la Tabla 4.1.

La única acción considerada es el peso propio, expresado como fuerza por unidad de volumen, $f_z = -\rho g = -24\,525$ N/m^3, con $\rho = 2500$ kg/m^3 y $g = 9{,}81$ m/s^2. La coherencia dimensional puede comprobarse en los propios datos, ya que cada muestra incluye los elementos discretos de superficie ds y de volumen dv, que cumplen exactamente $dv = h\,ds$ con $h = 0{,}1$ m, de modo que el producto $f_z\,dv = \rho g\,h\,ds$ recupera el peso propio de cada elemento de lámina y las integrales de trabajo (Sección 4.2) quedan en unidades de energía.

Todos los campos se interpolan a la rejilla regular de 64×64. En total se dispone de 6058 archivos, de los cuales 2400 son geometrías *solid*, sin hueco central, y 3658 son geometrías *hole*, con una región sin material en la que f_z, ds y dv se anulan. El estudio principal se realiza sobre el subconjunto *solid*, por ser el régimen geométricamente más homogéneo y, por tanto, el que aísla con más limpieza el comportamiento del surrogado y del guiado. El subconjunto *hole* se reserva como contraste de robustez y exige métricas enmascaradas.

Tabla 4.1: Definición del problema mecánico del conjunto de datos (Lourenço y cols., 2024).

Dominio en planta	10×10 m^2 (rejilla 64×64)
Espesor de la lámina	$h = 0{,}1$ m
Material	hormigón elástico lineal isótropo
Módulo de Young	$E = 30$ GPa
Coeficiente de Poisson	$\nu = 0{,}2$
Densidad	$\rho = 2500$ kg/m^3
Carga	peso propio, $f_z = -\rho g = -24\,525$ N/m^3
Apoyos	empotramiento en las esquinas
Solver	Abaqus/Standard, estático lineal, elementos S4R/S3R ($\approx 0{,}25$ m)

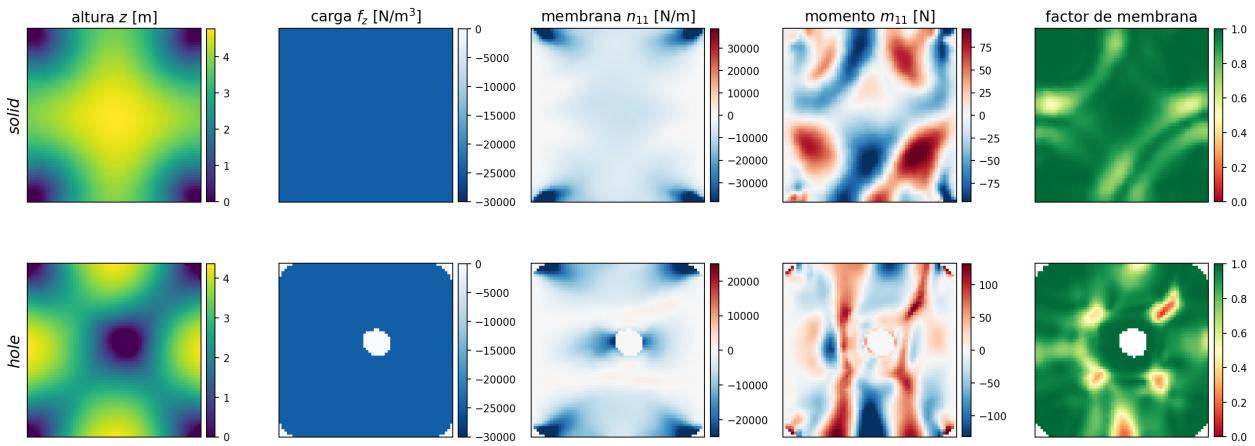

Figura 4.1: Una muestra de cada régimen del conjunto de datos (arriba *solid*, abajo *hole*). De izquierda a derecha, el mapa de alturas z, la carga de peso propio f_z (uniforme sobre el material y nula en el hueco), un esfuerzo de membrana representativo (n_{11}), suave y concentrado cerca de los apoyos, un momento flector representativo (m_{11}), más localizado y de alta frecuencia espacial, y el mapa de factor de membrana, próximo a uno donde domina la acción de membrana. En *hole*, la región sin material queda excluida de la métrica (en blanco).

La Figura 4.1 ilustra una muestra de cada régimen con sus campos principales. Cada archivo incluye la geometría $z \in \mathbb{R}^{1\times64\times64}$, la carga vertical $f_z \in \mathbb{R}^{1\times64\times64}$, la respuesta mecánica $y \in \mathbb{R}^{13\times64\times64}$, los elementos discretos *ds* y *dv* que se usan como pesos de cuadratura y máscaras, y el mapa de factor de membrana real calculado a partir de la respuesta FEM.

La respuesta mecánica sigue el orden canónico fijado en el código del proyecto, recogido en la Tabla 4.2. El canal de desplazamiento se separa de los campos de membrana y flexión porque su escala y su estructura espacial son distintas.

Tabla 4.2: Canales mecánicos predichos por el Ingeniero.

Bloque	Canales	Interpretación
Desplazamiento	u_z	desplazamiento vertical
Membrana	$\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{12}$	deformaciones de membrana
Membrana	n_{11}, n_{22}, n_{12}	esfuerzos de membrana
Flexión	$\kappa_{11}, \kappa_{22}, \kappa_{12}$	variaciones de curvatura
Flexión	m_{11}, m_{22}, m_{12}	momentos flectores

La partición se realiza con semilla fija 42 y proporción 80/20 entre entrenamiento y validación, dentro del régimen que se esté estudiando. Para *solid*, esto da 1920 muestras de entrenamiento y 480 de validación, y para *hole*, 2926 y 732 respectivamente.

La normalización se calcula exclusivamente sobre entrenamiento, evitando fuga de información. La geometría y la carga se llevan a $[-1, 1]$ mediante los mínimos y máximos del conjunto de entrenamiento,

$$z_{\text{norm}} = 2\frac{z - z_{\text{mín}}}{z_{\text{máx}} - z_{\text{mín}}} - 1, \qquad f_{z,\text{norm}} = 2\frac{f_z - f_{z,\text{mín}}}{f_{z,\text{máx}} - f_{z,\text{mín}}} - 1. \tag{4.1}$$

Los 13 canales físicos se estandarizan canal a canal,

$$y^{(c)}_{\text{norm}} = \frac{y^{(c)} - \mu_c}{\sigma_c}, \tag{4.2}$$

con μ_c y σ_c estimados solo en entrenamiento. Para calcular energías, factor de membrana y residuos físicos, las predicciones se desnormalizan de nuevo al espacio físico.

En el régimen *hole*, las regiones sin material tienen $ds = 0$ y campos mecánicos degenerados. Para evitar que esos ceros dominen la pérdida o las métricas, se define una máscara de material

$$M(i, j) = \mathbb{1}[ds(i, j) > 0], \tag{4.3}$$

y las pérdidas supervisadas y las métricas estructurales se promedian solo sobre $M = 1$. Esta decisión no cambia el problema físico, ya que simplemente evita evaluar en las regiones sin material y, por tanto, no hay respuesta mecánica que evaluar.

2. El pipeline: un prior de difusión y un surrogado mecánico

La propuesta se organiza en dos redes que se entrenan por separado y solo se acoplan en inferencia. El **Arquitecto** es un prior geométrico de difusión que aprende a generar mapas de altura plausibles. El **Ingeniero** es un surrogado diferenciable que, dada una geometría (posiblemente ruidosa) y una carga, estima su respuesta mecánica de 13 canales. Durante la generación, el gradiente del factor de membrana calculado a través del Ingeniero empuja la trayectoria del Arquitecto hacia geometrías mecánicamente más eficientes. El surrogado sustituye así al solver dentro del bucle de generación.

Esta separación, cuya motivación se expuso en la Introducción, es el rasgo de diseño más característico del pipeline. Las dos redes solo se comunican a través de un gradiente diferenciable en inferencia, de modo que cada componente se entrena, se valida y se sustituye de forma independiente, y el co-

nocimiento físico se inyecta sin reentrenar el prior. Es precisamente esta modularidad la que permite que el estudio del Ingeniero y del guiado, objeto de esta memoria, se realice sobre un Arquitecto fijo. La Figura 4.2 resume esta organización, separando el entrenamiento independiente de ambas redes del muestreo guiado en el que cooperan.

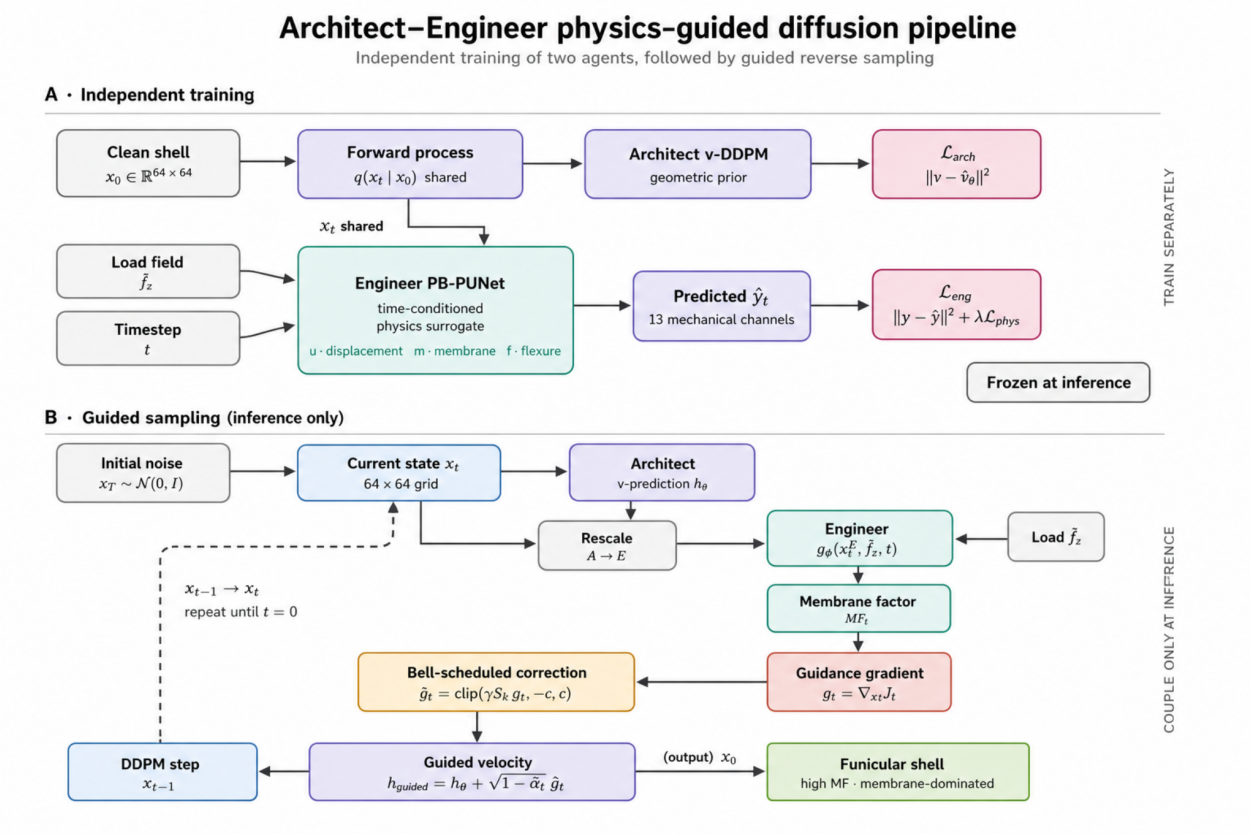

Figura 4.2: Esquema del pipeline Arquitecto–Ingeniero. Arriba, el entrenamiento independiente: el Arquitecto aprende el prior de difusión con una pérdida de v-prediction, y el Ingeniero aprende la respuesta mecánica sobre estados ruidosos con una pérdida supervisada más un residuo físico. Abajo, el muestreo guiado, donde ambos se acoplan: en cada paso, el gradiente del factor de membrana calculado a través del Ingeniero corrige la velocidad del Arquitecto según un schedule temporal.

2.1. El prior de difusión (Arquitecto)

El prior geométrico es una red U-Net (Ronneberger y cols., 2015) de difusión que recibe el estado ruidoso x_t y el timestep t, y predice la velocidad de difusión bajo parametrización v. Es deliberadamente incondicional, de modo que no recibe la carga f_z ni una etiqueta de tipología. Esta elección separa el prior de formas de la física y deja todo el control mecánico al guiado en inferencia, que es donde se concentra la aportación del trabajo. Internamente sigue la topología habitual de las U-Net

de difusión, un *encoder-decoder* con bloques residuales, atención en las resoluciones intermedias y conexiones de salto, en el que el embedding del timestep se inyecta en cada bloque.

Aunque en esta memoria el prior se mantenga congelado, su entrenamiento forma parte de la propuesta. La red se entrena minimizando el error cuadrático entre la velocidad real de difusión y la que predice,

$$\mathcal{L}_{\text{arch}}(\theta) = \mathbb{E}_{x_0,\varepsilon,t}\big[\, \|v_t - h_\theta(x_t, t)\|^2 \,\big], \tag{4.4}$$

con la velocidad v_t definida en el capítulo anterior y el estado ruidoso x_t generado por el proceso directo. Ese proceso emplea un schedule de ruido coseno en lugar del lineal habitual, porque el coseno reparte la pérdida de señal de forma más gradual y conserva durante más tiempo la estructura de baja frecuencia de los mapas de altura, que es la información mecánicamente relevante (la comparación de ambos schedules se mostró en la Figura 3.3 del capítulo anterior).

El proceso de difusión se discretiza en $T = 1000$ pasos. El entrenamiento utiliza el optimizador AdamW (Loshchilov y Hutter, 2019) con una tasa de aprendizaje de $2 \cdot 10^{-5}$, decaimiento de pesos de 10^{-4} y gradient clip, con tamaño de lote 32 y semilla fija para garantizar la reproducibilidad. El presupuesto máximo es de 75 épocas con *early stopping* por pérdida de validación (paciencia de 15 épocas) y reducción de la tasa de aprendizaje en meseta. Por eso los dos priors no consumen el mismo número de épocas, ya que el de *solid* se detiene en torno a la época 50 y el de *hole* agota casi el presupuesto, como se verá en sus curvas de entrenamiento. Para los contrastes en el régimen con hueco se entrena un prior análogo sobre ese subconjunto, también incondicional. A partir de aquí, el prior actúa como un generador fijo y reproducible.

2.2. El Ingeniero consciente del ruido

El Ingeniero recibe el estado ruidoso y la carga normalizada concatenados como dos canales de entrada,

$$\xi_t = \text{concat}(x_t, f_{z,\text{norm}}), \tag{4.5}$$

junto con el timestep t, y produce una respuesta mecánica de 13 canales,

$$g_\phi(\xi_t, t) = \hat{y} \in \mathbb{R}^{13\times 64\times 64}. \tag{4.6}$$

La clave del diseño es **cómo** se entrena. Durante el entrenamiento, el estado ruidoso se genera con el mismo proceso directo que usa el prior,

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I), \tag{4.7}$$

con t muestreado de forma uniforme. Así, el surrogado ve exactamente la misma clase de estados ruidosos que encontrará durante el muestreo inverso, y no se le obliga nunca a operar sobre geometrías limpias fuera de su dominio. Como se justifica en el capítulo anterior, esto orienta la predicción hacia la respuesta media compatible con cada estado ruidoso y es lo que mitiga la brecha de Jensen al evaluar la física sobre x_t.

Frente a las alternativas basadas en reconstruir primero una geometría limpia o en aplicar directamente un surrogado limpio sobre x_t, el Ingeniero consciente del ruido evalúa cada estado en las condiciones para las que ha sido entrenado. De este modo, evita tanto el sesgo asociado a la no linealidad mecánica como el uso del modelo fuera de su distribución de entrenamiento, y proporciona una señal de guiado más fiable en los niveles intermedios de ruido, donde se define la forma global de la geometría.

3. Arquitecturas del Ingeniero

Para comparar los surrogados de forma justa, todas las condiciones experimentales se mantienen constantes y únicamente se modifica la arquitectura. En particular, se emplean idénticos datos, partición, semilla, schedule de ruido, número de épocas, optimizador y funciones de pérdida. El estudio incluye un baseline convolucional, PB-PUNet, y dos operadores espectrales, EquiNO y EquiNO+σ^2. Esta última variante es la única que estima la incertidumbre y constituye la base de la calibración y del guiado adaptativo posteriores.

3.1. PB-PUNet: baseline convolucional

PB-PUNet es la arquitectura de referencia, tomada del trabajo previo de Lourenço et al. (Lourenço y cols., 2024). Es una red paralela formada por tres U-Net convolucionales independientes que comparten la misma entrada ξ_t y el mismo embedding temporal. Cada U-Net se especializa en un

bloque físico, de forma que una predice el desplazamiento u_z, otra los seis canales de membrana y otra los seis de flexión. Sus salidas se concatenan para reconstruir los 13 canales. Internamente, cada U-Net es un *encoder-decoder* con submuestreo y sobremuestreo progresivos y conexiones *skip* entre niveles equivalentes, de modo que combina contexto global (en los niveles profundos) con detalle local (en los niveles superficiales).

La virtud de esta familia es que las convoluciones multiescala capturan bien bordes, concentraciones locales de tensión y patrones de alta frecuencia. Su inconveniente es el coste, ya que al replicar tres U-Net completas el modelo acumula del orden de 361 millones de parámetros y es el más lento en inferencia. Es el surrogado de referencia frente al que se mide si una parametrización espectral, mucho más ligera, basta para el régimen de información que deja la difusión.

El calificativo *Physics-Based* de su nombre proviene de su función de pérdida, y conviene detenerse en ella porque fija la receta de entrenamiento de todo el estudio. El entrenamiento combina el ajuste de los 13 campos físicos en relación con la respuesta del solver de elementos finitos con un término de pérdida física. Este término penaliza las desviaciones respecto al equilibrio mecánico a partir de los campos de membrana y flexión predichos. La función de pérdida queda definida como $\mathcal{L}_{\text{sup}} + \lambda_{\text{phys}}\mathcal{L}_{\text{phys}}$, por lo que el modelo no solo aproxima la solución del solver, sino que también favorece predicciones físicamente coherentes. Esta formulación, heredada de PB-PUNet, se aplica sin cambios a los operadores espectrales, garantizando que las diferencias observadas se deban a la arquitectura y no a la función de pérdida. La Figura 4.3 esquematiza la arquitectura.

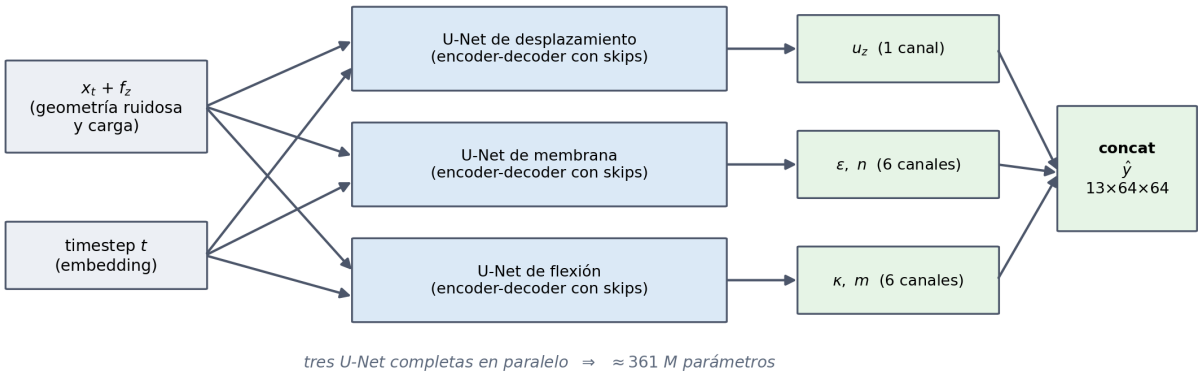

Figura 4.3: Esquema de PB-PUNet, el baseline convolucional (Lourenço y cols., 2024). Tres U-Net completas trabajan en paralelo sobre la misma entrada (estado ruidoso, carga y embedding del timestep), cada una especializada en un bloque físico (desplazamiento, membrana y flexión), y sus salidas se concatenan en los 13 canales de la respuesta.

3.2. EquiNO: operador espectral

EquiNO es un operador neuronal basado en las *Fourier Neural Operators*. A diferencia de las redes convolucionales, que aprenden filtros locales, este modelo transforma y procesa la información en el dominio de la frecuencia. En primer lugar, la entrada se combina con las coordenadas normalizadas de la rejilla y se proyecta a una representación latente de mayor dimensión. Estas coordenadas permiten al modelo distinguir la posición de cada punto y evitan que trate de igual forma patrones idénticos situados en regiones diferentes. Después, el campo atraviesa una pila de bloques de operador, y cada bloque combina en paralelo tres ingredientes.

- Una **convolución espectral**, que aplica la FFT 2D, conserva solo los primeros modos de Fourier en cada dirección, los multiplica por pesos complejos aprendidos y reconstruye el campo con la FFT inversa. Al truncar los modos altos, el bloque tiene un campo receptivo global con muy pocos parámetros y filtra de raíz el ruido de alta frecuencia.
- Una **componente puntual** 1×1, que actúa como una mezcla local de canales y aporta la parte de alta frecuencia que la rama espectral descarta.
- Un **embedding temporal** sinusoidal del timestep t, que se inyecta en cada bloque para que el operador module su comportamiento según el nivel de ruido.

Tras la pila de bloques, la salida se decodifica en tres ramas u, m y f, cada una con cabezas locales

y un término modal de bajo rango que captura de forma compacta los patrones globales dominantes de cada bloque. En los experimentos principales se usan 16 modos espectrales en cada dirección, ancho de operador 128, seis capas de operador, embedding temporal de dimensión 256, rango modal 12, peso residual modal 0.25 y dropout 0.05.

Para fijar el funcionamiento interno conviene seguir las formas de los tensores. Tras la proyección de entrada, el campo que circula por la red es un bloque de $128 \times 64 \times 64$, de modo que en cada posición espacial (un píxel) hay un vector de 128 características, y cada uno de los 128 canales es una imagen de 64×64. La convolución espectral actúa sobre las dos dimensiones espaciales de cada canal. La transformada rápida de Fourier de señal real lleva cada imagen 64×64 a un espectro de 64×33 coeficientes complejos, es decir, 2112 modos por canal, cada uno una onda espacial con su amplitud y su fase. De esos modos el operador conserva únicamente los 512 de frecuencia más baja y descarta el resto, un *low-pass filter* aprendido alineado con la información que sobrevive al ruido. En cada modo conservado, un peso aprendido con forma de matriz compleja 128×128 recombina los canales, lo que supone del orden de 8,4 millones de parámetros por convolución espectral, y multiplicado por las seis capas estos pesos concentran la mayor parte de los ≈ 102 millones de parámetros del modelo. La transformada inversa devuelve el campo a $128 \times 64 \times 64$.

La decodificación final se apoya en convoluciones 1×1, que equivalen a multiplicar, en cada píxel y de forma independiente, el vector de canales por una matriz, y así cambian el número de canales sin tocar la rejilla espacial. Tres cabezas de rama, una por bloque físico, llevan los 128 canales a 1, 6 y 6, que se concatenan en los 13 de la salida. Cada cabeza suma una predicción local píxel a píxel y una componente modal de bajo rango, una combinación de unos pocos patrones globales aprendidos que captura de forma barata la estructura dominante de cada campo. La Tabla 4.3 resume este recorrido.

Tabla 4.3: Recorrido de formas a través de EquiNO+σ^2.

Etapa	Operación	Forma de salida
entrada	x_t, f_z + coordenadas	$4 \times 64 \times 64$
proyección	convolución 1×1, $4 \to 128$	$128 \times 64 \times 64$
$6\times$ bloque	espectral + puntual + tiempo	$128 \times 64 \times 64$
salida del operador	norma + convolución 1×1	$128 \times 64 \times 64$
cabezas $u/m/f$	convolución 1×1 a $1, 6, 6$ + concatenar	$13 \times 64 \times 64$ (μ)
cabeza de incertidumbre	convolución 1×1 a 1	$1 \times 64 \times 64$ ($\log \sigma^2$)

El objetivo de EquiNO no es aumentar la capacidad del modelo, sino comprobar si una representación espectral centrada en las bajas frecuencias resulta más adecuada para la información que permanece reconocible durante la difusión. Con esta configuración, EquiNO cuenta con aproximadamente 102 millones de parámetros, unas 3,5 veces menos que PB-PUNet, y reduce cerca de la mitad el tiempo de inferencia. La Figura 4.4 muestra la arquitectura completa y la estructura interna de cada bloque del operador.

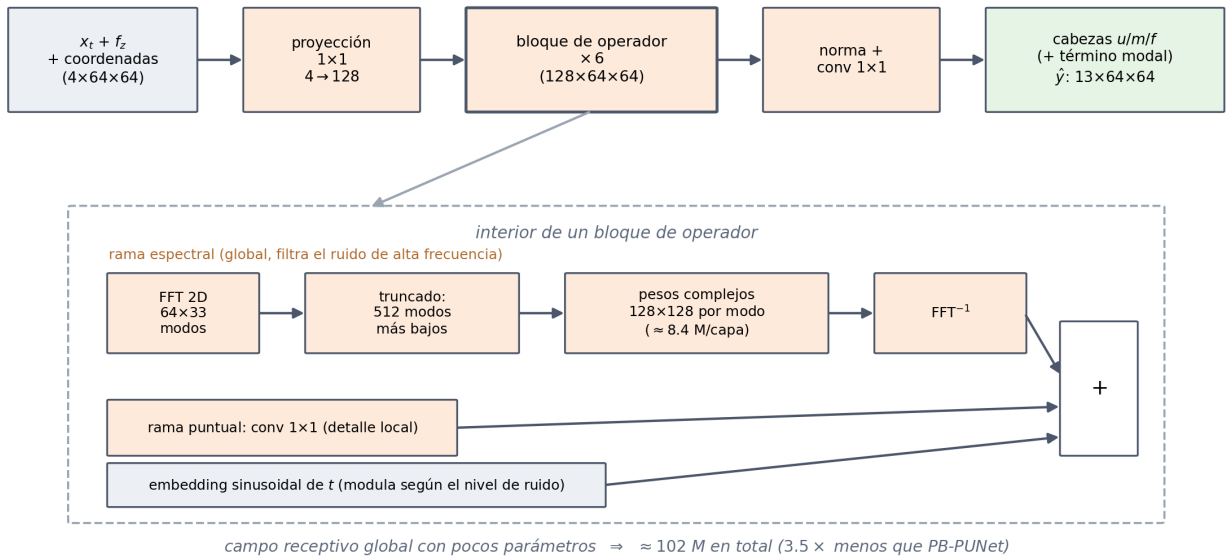

Figura 4.4: Esquema de EquiNO. Arriba, el flujo completo: proyección 1×1, seis bloques de operador y las cabezas de salida por rama. Abajo, el interior de un bloque: la rama espectral (FFT, truncado a los 512 modos más bajos, pesos complejos por modo y FFT inversa) aporta el campo receptivo global y filtra el ruido de alta frecuencia, la rama puntual 1×1 aporta el detalle local, y el embedding del timestep modula el bloque según el nivel de ruido.

3.3. EquiNO+σ^2: la cabeza de incertidumbre

EquiNO+σ^2 es exactamente el operador anterior con una única adición. Junto a la media μ_ϕ, una segunda cabeza predice un mapa de log-varianza

$$\ell_\phi(x_t, t) = \log \sigma^2_\phi(x_t, t), \tag{4.8}$$

clampado en $[-6, 3]$ durante el entrenamiento por estabilidad numérica. El motivo de acotar la log-varianza está en la propia forma de la verosimilitud heterocedástica, cuyo término de ajuste

es $\exp(-\ell_\phi)(y-\mu_\phi)^2$. Si la red pudiera llevar ℓ_ϕ a valores muy negativos, es decir, declarar una varianza casi nula, el factor de precisión $\exp(-\ell_\phi)$ crecería sin control y amplificaría cualquier residuo hasta desbordar el gradiente. La cota inferior $\ell_\phi \geq -6$ lo evita, ya que limita la precisión máxima a $\exp(6) \approx 400$. En el otro extremo, si la red pudiera llevar ℓ_ϕ a valores muy grandes, podría minimizar la pérdida declarando una incertidumbre enorme, con lo que $\exp(-\ell_\phi) \to 0$ apagaría la señal que ajusta la media. La cota superior $\ell_\phi \leq 3$ lo impide y mantiene vivo el ajuste. Ambos límites se fijan en el espacio normalizado, donde los campos tienen varianza del orden de la unidad, de modo que el intervalo $[\exp(-6), \exp(3)] \approx [0{,}002, 20]$ cubre con holgura el rango plausible del error por píxel a la vez que descarta los dos modos de fallo numérico. La varianza se predice como un único mapa espacial que se expande sobre los 13 canales para evaluar la verosimilitud heterocedástica. La columna espectral, los hiperparámetros y el conteo de parámetros son idénticos a EquiNO, y solo cambia esa salida adicional. Por eso la pareja EquiNO frente a EquiNO+σ^2 aísla limpiamente el efecto de añadir incertidumbre, sin confundirlo con un cambio de arquitectura.

La misma red se entrena por separado en *solid* y en *hole*, cambiando únicamente el subconjunto de datos y, en *hole*, el enmascarado de la pérdida por material. Esta cabeza de σ^2 es la que habilita después la calibración de la incertidumbre y el guiado adaptativo, ya que el schedule temporal de guiado se deriva directamente de la fiabilidad estimada por la red. La Figura 4.5 muestra la estructura de doble salida.

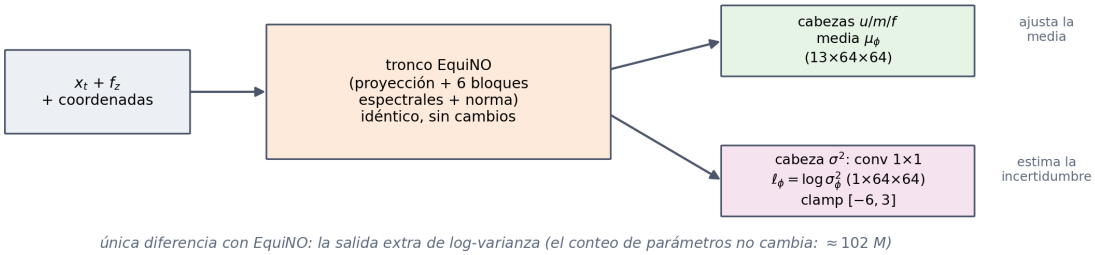

Figura 4.5: Esquema de EquiNO+σ^2. El tronco espectral es idéntico al de EquiNO, y la única diferencia es la segunda salida, una cabeza 1×1 que predice el mapa de log-varianza $\ell_\phi = \log\sigma^2_\phi$, con el que la red declara su propia fiabilidad en cada píxel y timestep.

La Tabla 4.4 resume la configuración y el coste de las tres arquitecturas, quedando así explícito qué cambia entre cada par de modelos comparados.

Tabla 4.4: Configuración y coste de las tres arquitecturas de Ingeniero comparadas. El conteo de parámetros y el tiempo de inferencia se miden en el benchmark del Capítulo de resultados.

Modelo	Tipo	Rasgos clave	Parámetros
PB-PUNet	3 U-Net paralelas (convolucional)	ramas $u/m/f$, convolución multiescala, sin σ^2	≈ 361 M
EquiNO	operador espectral (FNO)	convolución de Fourier + puntual, ramas $u/m/f$ + modal	≈ 102 M
EquiNO+σ^2	operador espectral + cabeza heterocedástica	igual que EquiNO más mapa de log-varianza	≈ 102 M

4. Entrenamiento del Ingeniero

Las tres familias se entrenan bajo las mismas condiciones, de modo que la comparación refleje únicamente las diferencias entre arquitecturas. La pérdida combina un término de ajuste a los campos FEM, un término de incertidumbre (solo cuando hay cabeza de σ^2) y una regularización física ligera.

4.1. Ajuste a los campos y verosimilitud heterocedástica

El término básico es una regresión de los trece campos, en la que la red debe acercar su predicción $\hat{y}$ a la respuesta FEM y, ambas en espacio normalizado. En el régimen *solid* no interviene ninguna máscara, y la pérdida es simplemente el error cuadrático medio sobre canales y píxeles,

$$\mathcal{L}_{\text{sup}} = \frac{1}{C\,HW} \left\| \hat{y}_{\text{norm}} - y_{\text{norm}} \right\|^2, \tag{4.9}$$

con $\|\cdot\|$ la norma euclídea sobre los $C = 13$ canales y los $H \times W$ píxeles.

La máscara solo aparece cuando procede, es decir, en el régimen *hole*, donde la geometría contiene una región sin material en la que la respuesta mecánica no está definida, y penalizar a la red en ella sería exigirle acertar valores sin significado físico. En ese caso el mismo error cuadrático se promedia únicamente sobre el material, mediante una máscara binaria M que vale $M(i,j) = 1$ donde hay material y 0 en el hueco,

$$\mathcal{L}_{\text{sup}} = \left\langle (\hat{y}_{\text{norm}} - y_{\text{norm}})^2 \right\rangle_M, \qquad \langle A \rangle_M = \frac{\sum_{c,i,j} M(i,j)\, A(c,i,j)}{C \sum_{i,j} M(i,j) + \epsilon}. \tag{4.10}$$

Ambas expresiones coinciden cuando $M \equiv 1$, de modo que (4.10) generaliza a (4.9) y en el resto del

capítulo se usa la notación $\langle\cdot\rangle_M$ con este convenio, máscara de material en *hole* y dominio completo en *solid*. Se registran además pérdidas por bloque (desplazamiento, membrana y flexión) que no intervienen en la optimización, pero permiten diagnosticar qué parte de la física domina el error.

Cuando el modelo lleva cabeza de incertidumbre y predice la log-varianza $\ell_\phi = \log \sigma_\phi^2$, la pérdida supervisada se sustituye por la verosimilitud negativa gaussiana heterocedástica,

$$\mathcal{L}_{\mathrm{NLL}} = \frac{1}{2} \left\langle \exp(-\ell_\phi)(y_{\mathrm{norm}} - \mu_\phi)^2 + \ell_\phi \right\rangle_M , \tag{4.11}$$

con la misma convención de máscara. Frente al error cuadrático, esta pérdida obliga a la red a explicar a la vez la media del error (a través de μ_ϕ) y su escala (a través de ℓ_ϕ). Si declara poca varianza donde falla, el término $\exp(-\ell_\phi)$ amplifica la penalización, y si declara varianza de más, el término ℓ_ϕ la castiga. En esta memoria, σ_ϕ^2 se interpreta como la incertidumbre aleatoria asociada a la posterior que induce el estado ruidoso, ya que en el óptimo la cabeza recupera la varianza del error condicional, que es la misma frontera de información que estudia el Marco Teórico.

4.2. Regularización física

Para anclar la predicción a un balance mecánico aproximado, se añade un residuo de trabajo virtual construido con los propios campos predichos. A partir de los canales de membrana y flexión se calculan las contribuciones elementales

$$\delta W_m = (n_{11}\varepsilon_{11} + n_{22}\varepsilon_{22} + 2n_{12}\varepsilon_{12})\, ds, \tag{4.12}$$

$$\delta W_f = (m_{11}\kappa_{11} + m_{22}\kappa_{22} + 2m_{12}\kappa_{12})\, ds, \qquad \delta W_{ext} = f_z u_z\, dv, \tag{4.13}$$

donde ds y dv son los pesos discretos de integración del dataset. Con f_z expresada como fuerza por unidad de volumen (Tabla 4.1), los tres términos quedan en unidades de energía y el balance es dimensionalmente consistente. El residuo global y su penalización son

$$\Delta P = \sum_{i,j} (\delta W_m + \delta W_f - \delta W_{ext})_{i,j}, \qquad \mathcal{L}_{\mathrm{phys}} = (\Delta P)^2. \tag{4.14}$$

Conviene delimitar su alcance, ya que al ser escalar este residuo puede satisfacerse por cancelación global aunque existan errores locales, y omite factores constantes, por lo que actúa como un regularizador operacional de trabajo virtual y no como una certificación dimensional. Se pondera con un *warmup* por época y con un peso dependiente del timestep

$$\omega_{\text{phys}}(t) = \left(1 - \frac{t}{T}\right)^{p}, \qquad p = 2, \tag{4.15}$$

de modo que pesa más cerca de los estados limpios, donde la interpretación mecánica de la geometría es fiable, y se atenúa en estados de alto ruido. La pérdida total del Ingeniero es entonces

$$\mathcal{L}_{\text{eng}} = \mathcal{L}_{\text{sup/NLL}} + \lambda_{\text{phys}}(e)\, \mathbb{E}[\omega_{\text{phys}}(t)\, \mathcal{L}_{\text{phys}}]\,. \tag{4.16}$$

La Tabla 4.5 recoge los hiperparámetros de entrenamiento, compartidos por las tres familias para que la comparación sea limpia.

Tabla 4.5: Hiperparámetros principales de entrenamiento del Ingeniero (comunes a las tres familias).

Bloque	Parámetro	Valor
Datos	split / semilla	80/20, seed 42
Datos	batch	8
Difusión	timesteps de entrenamiento	$T = 1000$, muestreo uniforme de t
Optimizador	AdamW	lr 10^{-4}, weight decay 10^{-4}
Entrenamiento	épocas / early stopping	60 / paciencia 15
LR scheduler	ReduceLROnPlateau	paciencia 8, factor 0.5
Estabilidad	grad clip / dropout	1.0 / 0.05
Física	$\lambda_{\text{phys}}^{\text{máx}}$ / warmup	10^{-3} / 10 épocas
Física	$\omega_{\text{phys}}(t)$	$(1 - t/T)^2$
Incertidumbre	clamp ℓ_ϕ / warmup NLL	$[-6, 3]$ / 10 épocas

5. Guiado por factor de membrana

5.1. Objetivo de guiado

Durante el muestreo, el Ingeniero predice una respuesta mecánica para el estado actual, de la que se obtiene el mapa de factor de membrana y su promedio espacial. Para cada muestra b de un *batch*

de tamaño B se define el objetivo

$$J_b(x_t) = \left(1 - \overline{\mathrm{mf}}_b(x_t)\right)^2, \tag{4.17}$$

y el gradiente de guiado se obtiene por retropropagación a través del Ingeniero,

$$g_t = \nabla_{x_t} \sum_b J_b(x_t). \tag{4.18}$$

El gradiente se suma sobre las B muestras del *batch* y no se promedia, tal como refleja la ecuación anterior. Al sumar en lugar de promediar, el gradiente respecto de una muestra coincide con el de su propio objetivo, ya que solo J_b depende de $x_t^{(b)}$,

$$\nabla_{x_t^{(b)}} \sum_{b'=1}^{B} J_{b'}(x_t) = \nabla_{x_t^{(b)}} J_b(x_t), \tag{4.19}$$

y no queda dividido por B, de modo que la intensidad del guiado que recibe cada lámina no depende del tamaño del *batch*. El promedio se reserva únicamente para registrar y comparar el valor del objetivo,

$$\overline{J} = \frac{1}{B} \sum_{b=1}^{B} J_b(x_t). \tag{4.20}$$

El prior y el Ingeniero pueden usar estadísticas de normalización distintas, así que antes de evaluar el surrogado el estado x_t, expresado en la escala del prior, se lleva primero a su escala física con los mínimos y máximos del prior,

$$x_t^{\mathrm{fis}} = \tfrac{1}{2}(x_t + 1)\left(z^{\mathrm{P}}_{\mathrm{máx}} - z^{\mathrm{P}}_{\mathrm{mín}}\right) + z^{\mathrm{P}}_{\mathrm{mín}}, \tag{4.21}$$

y a continuación se renormaliza con los del Ingeniero,

$$x_t^{\mathrm{ing}} = 2\,\frac{x_t^{\mathrm{fis}} - z^{\mathrm{E}}_{\mathrm{mín}}}{z^{\mathrm{E}}_{\mathrm{máx}} - z^{\mathrm{E}}_{\mathrm{mín}}} - 1. \tag{4.22}$$

La composición de ambas es una única transformación afín de pendiente $a = \left(z^{\mathrm{P}}_{\mathrm{máx}} - z^{\mathrm{P}}_{\mathrm{mín}}\right)/\left(z^{\mathrm{E}}_{\mathrm{máx}} - z^{\mathrm{E}}_{\mathrm{mín}}\right)$, por lo que es diferenciable y el gradiente calculado por el Ingeniero se propaga hasta x_t sin

interrumpirse.

Durante la generación no existe una carga asociada a la muestra, así que la entrada de carga del Ingeniero se fija en un valor constante del rango de entrenamiento (el punto central del rango normalizado), idéntico para todas las estrategias comparadas y para todos los jueces. Esta elección no favorece a ninguna estrategia y, en régimen elástico lineal, es además poco influyente sobre la métrica, ya que al escalar uniformemente la carga todos los campos escalan con ella y el factor de membrana, que es un cociente de energías, permanece invariante.

5.2. Campana fija: el guiado base

El guiado de referencia modula el gradiente con una envolvente temporal gaussiana,

$$w_k^{\text{bell}} = w_{\text{máx}} \exp\left(-\frac{(r_k - \rho)^2}{2\sigma_{\text{bell}}^2}\right), \qquad r_k = \frac{k}{K-1}, \tag{4.23}$$

con k el índice del paso inverso y K el número de pasos. La campana concentra la corrección en el ruido intermedio. En cada paso, el gradiente se escala, se recorta y se inyecta en la velocidad predicha por el prior,

$$\tilde{g}_t = \text{clip}(\gamma\, w_k\, g_t, -c, c), \qquad \tilde{v}_t = v_\theta(x_t, t) + \sqrt{1-\bar{\alpha}_t}\,\tilde{g}_t, \tag{4.24}$$

y se aplica el paso estándar del scheduler para obtener x_{t-1}. La forma de esta campana (su pico ρ y su anchura σ_{bell}) es un hiperparámetro que hay que fijar a mano.

5.3. Schedule temporal derivado de la incertidumbre

La aportación metodológica central es sustituir esa campana fijada a mano por un schedule temporal que la propia red decide a través de su incertidumbre. Primero se estima un perfil de varianza media por timestep,

$$u_t = \mathbb{E}_{\Omega,c}\left[\sigma_\phi^2(x_t, t)\right], \tag{4.25}$$

sobre geometrías de validación ruidificadas con el mismo scheduler, donde Ω recorre las muestras del lote y los píxeles del dominio y c los canales. Es la misma media que define u_t en el capítulo anterior

(ecuación 3.59). Después se define un peso proporcional a la precisión estimada,

$$w_t^{\sigma} = C \frac{1}{u_t + \epsilon}, \tag{4.26}$$

y la constante C se fija para igualar el presupuesto total de guiado de la campana,

$$\sum_t w_t^{\sigma} = \sum_t w_t^{\text{bell}}. \tag{4.27}$$

Así, la comparación modifica únicamente la distribución temporal del *schedule*, manteniendo constante la corrección total aplicada. El guiado se concentra en los pasos donde el surrogado presenta mayor confianza y se reduce cuando su incertidumbre aumenta, sin necesidad de ajustar manualmente la posición ni la anchura del máximo. La Figura 4.6 compara ambas envolventes, deduciendo la del schedule del perfil de incertidumbre del Ingeniero ya entrenado.

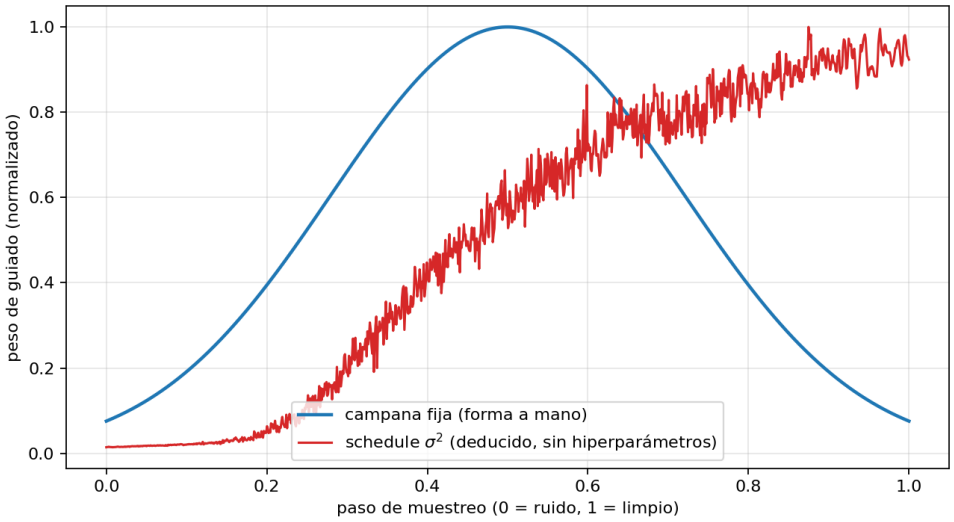

Figura 4.6: Envolvente temporal de guiado, medida con el Ingeniero entrenado. La campana fija impone una forma a mano que pica en mitad de la trayectoria, mientras que el schedule por incertidumbre deduce la suya del perfil de varianza u_t del propio Ingeniero, con el mismo presupuesto total, y concentra la corrección donde la red se estima más fiable, cerca de los estados limpios.

5.4. Guiado enmascarado en geometrías con hueco

En *hole*, optimizar el promedio global del factor de membrana podría premiar artefactos en regiones sin material. Para evitarlo, el objetivo se calcula solo sobre la máscara de material, estimada en cada

paso a partir de la geometría limpia implícita $\hat{x}_0$, de modo que el guiado no rellene el hueco para mejorar la métrica de forma artificial,

$$\overline{\mathrm{mf}}_M = \frac{\sum_{i,j} M(i,j)\,\mathrm{mf}(i,j)}{\sum_{i,j} M(i,j) + \epsilon}. \tag{4.28}$$

Esta variante solo se activa en los experimentos de *hole* y no altera la metodología sobre *solid*. La Tabla 4.6 reúne los hiperparámetros de generación y guiado empleados en todos los experimentos.

Tabla 4.6: Hiperparámetros de generación y guiado.

Bloque	Parámetro	Valor
Muestreo	pasos / seed	1000 / 7
Muestreo	muestras / batch	$N = 256$ / 8
Campana	$w_{\text{máx}}, \rho, \sigma_{\text{bell}}$	1,0, 0,5, 0,22
Gradiente	escala γ / recorte c	600 / 20
Schedule σ^2	geometrías para el perfil u_t	16 reales ruidificadas
Recalibración σ^2	temperatura τ (solid)	1,78
Bootstrap	réplicas B (según experimento)	2000–5000

6. Metodología de evaluación

La evaluación se organiza en tres bloques, que responden a tres preguntas, si el surrogado es preciso (R1), si su incertidumbre es fiable (R2) y si esa incertidumbre mejora la generación guiada (R3).

6.1. R1: benchmark de surrogados

Mide cómo se degrada cada Ingeniero a lo largo del espectro de ruido. Para cada muestra de validación y cada timestep de una rejilla

$$\mathcal{T}_{\text{bench}} = \{0, 50, 100, \ldots, 950, 999\}, \tag{4.29}$$

se genera un x_t con ruido pareado y se evalúa el surrogado. Las métricas principales son el error de campo y el error del factor de membrana,

$$\text{PhysMSE}(t) = \mathbb{E}\left[\|\hat{y} - y\|_2^2\right], \qquad \text{MF-MAE}(t) = \mathbb{E}\left[\left|\widehat{\mathrm{mf}} - \mathrm{mf}_{\text{true}}\right|\right], \tag{4.30}$$

y se registran además errores por rama, número de parámetros y tiempo de inferencia. En *hole*, todas las métricas de campo y de factor de membrana se calculan sobre la máscara de material.

Además del valor a lo largo de todo el espectro, se resume el error en la **ventana de guiado**, definida como el intervalo $t \in [280, 720]$ que corresponde a una desviación $\pm\sigma_{\text{bell}}$ alrededor del pico de la campana ($\rho = 0{,}5$, $\sigma_{\text{bell}} = 0{,}22$, sobre $T = 1000$). Es la zona de la trayectoria donde la corrección física se concentra, y por eso el error ponderado en esa ventana es la métrica más relevante para el guiado.

6.2. R2: calibración de la incertidumbre

Para el modelo con cabeza de varianza se evalúa si σ_ϕ^2 predice el error real. En espacio normalizado se define el error cuadrático medio por píxel

$$e^2(i,j) = \frac{1}{13}\sum_{c=1}^{13} \left(\hat{y}_c(i,j) - y_c(i,j)\right)^2 , \tag{4.31}$$

y se miden dos diagnósticos. El primero, la correlación entre σ_ϕ^2 y e^2 (global y por timestep), responde a si la varianza **ordena** bien el error. El segundo, un diagrama de fiabilidad que compara la varianza media predicha con el error real a lo largo del tiempo de difusión, responde a si su **escala** es correcta (Levi y cols., 2019). Cuando la varianza ordena bien el error pero su escala es incorrecta, se aplica una recalibración por temperatura $\tilde{\sigma}^2 = \tau\,\sigma_\phi^2$, con τ ajustado en una partición y evaluado en otra. Esta calibración se realiza antes de usar la incertidumbre como schedule, para separar la información de ranking de la escala probabilística.

6.3. R3: evaluación de la generación guiada

La generación se evalúa con el prior congelado y con semillas pareadas entre estrategias, comparando tres condiciones, sin guiado (prior puro), guiado con campana fija (base) y guiado con el mismo Ingeniero pero schedule derivado de σ^2. Se usa muestreo fiel de 1000 pasos en los resultados finales.

La calidad de las geometrías generadas se evalúa mediante el factor de membrana. En ambos regímenes, esta métrica se calcula con EquiNO+σ^2, el surrogado más preciso del estudio, cuya capacidad de estimación se valida previamente comparando sus predicciones con los valores FEM del conjunto de

validación. PB-PUNet se emplea únicamente como evaluador de contraste, ya que pertenece a una familia arquitectónica diferente y permite comprobar que las mejoras no dependen exclusivamente del modelo utilizado durante el guiado. Dado que las muestras generadas no se simulan de nuevo mediante FEM, los resultados deben interpretarse como una comparación relativa entre estrategias y no como una validación estructural definitiva.

La diversidad se mide sobre las geometrías generadas. En el espacio de píxel se usan estadísticas de dispersión. En el espacio latente se entrena un VAE auxiliar sobre geometrías reales del conjunto de entrenamiento y se calculan distancias par a par y, como métrica de diversidad de referencia, la cobertura de *convex hull* sobre una proyección PCA-2D del latente. Se elige el área de *convex hull* por coherencia con la práctica de la línea de trabajo de partida y por su lectura directa como volumen de exploración del diseño, y la proyección PCA-2D se mantiene fija porque cambia el área numérica. Esta forma de evaluar sigue la recomendación de juzgar generadores de ingeniería por factibilidad, novedad y diversidad, y no solo por similitud estadística agregada (Regenwetter, Srivastava, Gutfreund, y Ahmed, 2023).

El VAE auxiliar es una posible fuente de sesgo, por lo que se fija su configuración, con encoder convolucional $1 \to 32 \to 64 \to 128 \to 128$ con stride 2, latente de dimensión 16, decoder transpuesto simétrico, activación final tanh, Adam con tasa de aprendizaje 10^{-3}, lote 64, 40 épocas y peso KL $\beta = 10^{-3}$. Se entrena con el split de entrenamiento del régimen evaluado, normalizando con las estadísticas del prior.

6.4. Significancia y reproducibilidad

Se resumen aquí las decisiones de protocolo comunes a todos los experimentos. Los entrenamientos usan la partición 80/20 con semilla fija dentro del régimen correspondiente, y la generación usa una semilla común con desplazamientos deterministas por lote, de modo que todas las estrategias comparadas parten exactamente del mismo ruido inicial. El benchmark de surrogados se evalúa sobre la validación completa de cada régimen (480 muestras en *solid*, 732 en *hole*) con ruido pareado por timestep, y la generación con muestreo fiel de 1000 pasos y $N = 256$ muestras por estrategia. La recalibración de la incertidumbre ajusta la temperatura τ en una mitad de una submuestra de validación y la evalúa en la otra. Las comparaciones entre estrategias se cuantifican con bootstrap

pareado (entre 2000 y 5000 réplicas según el experimento), reportando intervalos de confianza al 95 %. Este bootstrap mide la incertidumbre de evaluación, no la de entrenamiento, pues cada arquitectura se entrena con una única semilla, de modo que las diferencias se leen condicionadas a esos entrenamientos. Cada resultado queda asociado a un checkpoint concreto, con su configuración, sus estadísticas de normalización y sus curvas archivadas para su reproducción. La Tabla 4.7 resume el diseño experimental completo, sirviendo de mapa para el capítulo de resultados.

Tabla 4.7: Resumen del diseño experimental.

Bloque	Pregunta	Variable principal	Métricas
R1	¿es preciso el surrogado?	familia de arquitectura	PhysMSE, MF-MAE, parámetros, tiempo
R2	¿es fiable su σ^2?	calibración / recalibración	correlación, diagrama de fiabilidad
R3	¿mejora el guiado?	schedule temporal	MF, convex hull, frontera mf–diversidad

En conjunto, esta metodología busca aislar las variables de interés, de modo que el prior se mantiene fijo, los splits y ruidos se parean, y las métricas distinguen precisión cruda, precisión del factor de membrana y utilidad para guiar. Esa separación es necesaria porque una arquitectura puede ser mejor en PhysMSE y peor en MF-MAE, o producir un gradiente menos útil aunque su error medio sea bajo. El objetivo no es elegir la red más grande, sino identificar cuándo el surrogado aporta una señal mecánica fiable para modificar la trayectoria de difusión.

El código completo del pipeline, las configuraciones de entrenamiento, los scripts de generación y de evaluación, y las utilidades para reproducir las figuras de esta memoria están disponibles de forma pública en el repositorio del proyecto.[1]

[1] `https://github.com/DanielArizaGarcia/Hybrid-PI-DDPM`

Capítulo 5

Resultados y discusión

Este capítulo presenta los resultados de forma gradual, siguiendo el orden en que se construye el sistema. El conjunto de la evaluación valida dos cosas a la vez, que el pipeline modular funciona como sistema, es decir, que un prior de difusión entrenado por separado y un surrogado mecánico acoplado solo en inferencia generan láminas eficientes de forma estable, y que sus componentes intervenibles admiten las mejoras que propone la memoria. Primero se muestra que el prior de difusión está bien entrenado y se congela. Después se compara el surrogado mecánico frente al baseline y se explica por qué un operador espectral resulta más eficiente, apoyándose en el régimen de información que deja la difusión y en la brecha de Jensen. A continuación se comprueba que la incertidumbre del surrogado es informativa y se calibra. Finalmente, una vez definidos y validados todos los componentes, el modelo se utiliza para generar láminas guiadas y evaluar tanto su calidad mecánica como la diversidad de las geometrías obtenidas. Todas las comparaciones se hacen con el prior congelado, con *splits* y ruidos pareados, y los resultados del régimen con hueco se presentan junto a los de *solid* donde procede. A lo largo de este capítulo, el factor de membrana se evalúa siempre con EquiNO+σ^2, tanto en el régimen *solid* como en *hole*. PB-PUNet solo se emplea, cuando se indica expresamente, como evaluador de contraste para comprobar que los resultados no dependen del mismo modelo utilizado durante el guiado.

1. El prior de difusión

El primer elemento del pipeline es el Arquitecto, el prior de difusión que genera la geometría. Se entrenan dos priors independientes, uno para el régimen *solid* y otro para el régimen *hole*, y ambos se congelan una vez convergen. La Figura 5.1 muestra sus curvas de entrenamiento, y la pérdida de difusión cae de forma estable y las curvas de entrenamiento y validación se mantienen juntas, sin signos de sobreajuste, en los dos regímenes.

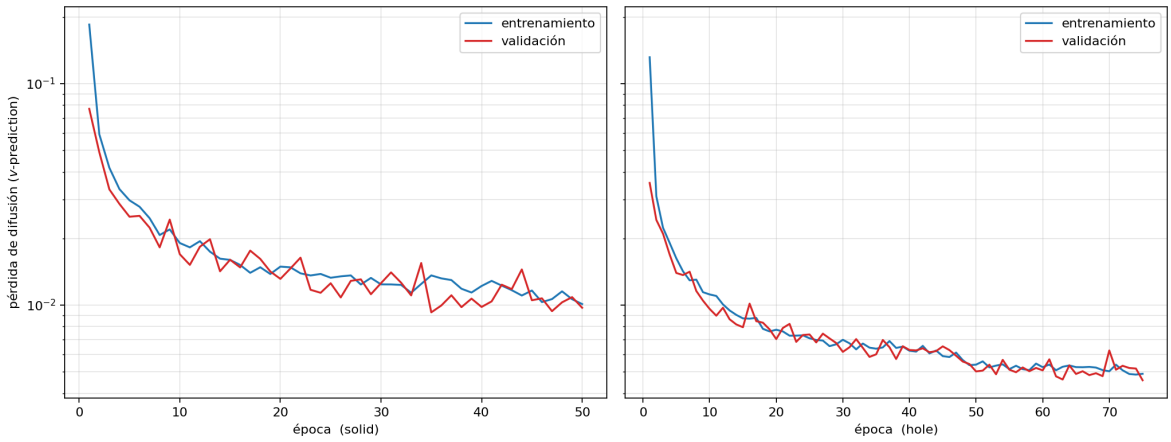

Figura 5.1: Curvas de entrenamiento del prior de difusión para *solid* (izquierda) y *hole* (derecha). La pérdida de v-prediction desciende de forma estable y la validación sigue al entrenamiento, por lo que el prior se considera convergido y se congela para el resto del trabajo.

A partir de este punto, el prior se mantiene fijo y se utiliza como generador reproducible de láminas funiculares. El resto del capítulo se centra en el Ingeniero y en el uso de su incertidumbre para construir una señal de guiado fiable.

2. El surrogado mecánico: comparación de arquitecturas

La primera aportación experimental es arquitectónica. Se comparan tres Ingenieros, el baseline convolucional PB-PUNet, el operador espectral EquiNO y su variante con incertidumbre EquiNO+σ^2, sobre el mismo conjunto de validación y con ruido pareado a lo largo de todo el espectro de difusión. La métrica relevante para el guiado es el error en el factor de membrana, sobre todo en la ventana de ruido intermedio donde el guiado actúa. La Figura 5.2 muestra la degradación de los tres modelos a lo largo del ruido, tanto en el factor de membrana como en el error de campo.

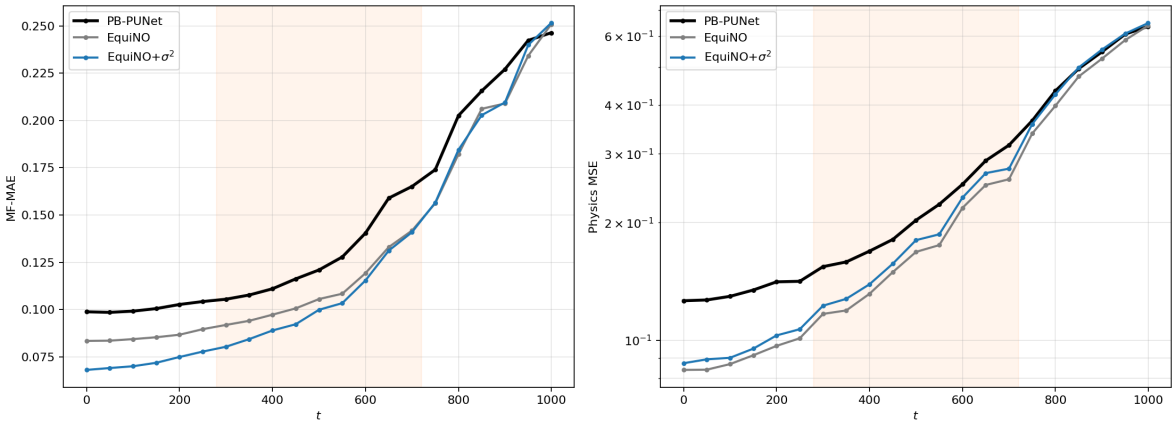

Figura 5.2: Degradación de cada surrogado a lo largo del ruido en *solid*. Izquierda: error en el factor de membrana, MF-MAE(t). Derecha: error de campo, PhysMSE(t) en escala logarítmica. La banda sombreada marca la ventana de guiado. Los dos operadores espectrales quedan claramente por debajo del baseline convolucional en MF-MAE, y EquiNO+σ^2 es el mejor de los tres.

La Tabla 5.1 resume los resultados. En la ventana de guiado, EquiNO+σ^2 reduce el error del factor de membrana un 17,2 % respecto a PB-PUNet y un 5,2 % frente a EquiNO sin incertidumbre. Esta mejora es especialmente relevante porque el guiado se aplica sobre estados intermedios y ruidosos, no únicamente sobre la geometría limpia final. Por tanto, una mayor precisión en esta región permite obtener una señal mecánica más fiable para modificar la trayectoria de difusión, incluso cuando las arquitecturas presentan errores similares al final del proceso. Además, EquiNO+σ^2 logra estos resultados con aproximadamente 3,5 veces menos parámetros y cerca de la mitad del tiempo de inferencia.

Tabla 5.1: Comparación de surrogados en *solid* (validación, ruido pareado). MF-MAE@t_0 es el error del factor de membrana en estado limpio; la columna ponderada usa la ventana de guiado.

Modelo	MF-MAE@t_0	MF-MAE (ventana)	PhysMSE@t_0	Parámetros	vs PB-PUNet
EquiNO+σ^2	**0.068**	**0.112**	0.087	≈ 102 M	−17,2 %
EquiNO	0.083	0.118	0.084	≈ 102 M	−12,7 %
PB-PUNet	0.099	0.135	0.126	≈ 361 M	—

2.1. Por qué se gana en factor de membrana y no en error de campo

Hay un matiz que conviene explicar bien, porque a primera vista parece contradictorio. EquiNO+σ^2 gana en el factor de membrana, pero en el error de campo crudo (PhysMSE@t_0) queda ligeramente por encima de EquiNO (0,087 frente a 0,084). ¿Cómo puede un modelo predecir mejor el factor de

membrana si este se calcula a partir de los mismos campos físicos?

La clave es que el factor de membrana no es una magnitud absoluta, sino un **cociente**, ya que mide qué fracción del trabajo estructural se conduce por membrana frente a flexión, $\mathrm{mf} = W_m/(W_m+W_f)$. El PhysMSE penaliza el error absoluto de cada canal, mientras que el factor de membrana solo depende del **balance** entre membrana y flexión. Un ejemplo lo deja claro. Supóngase que el reparto real es $W_m = 9$, $W_f = 1$, de modo que $\mathrm{mf} = 0{,}90$. Un modelo que sobrestime ambos términos por igual, prediciendo $W_m = 18$, $W_f = 2$, tiene un PhysMSE enorme y, sin embargo, un factor de membrana casi perfecto ($\mathrm{mf} = 0{,}90$). Al revés, un modelo con errores absolutos pequeños pero que desplace parte del trabajo de membrana hacia flexión, prediciendo $W_m = 8{,}5$, $W_f = 1{,}5$, tiene un PhysMSE bajo pero un factor de membrana sesgado ($\mathrm{mf} = 0{,}85$). Las dos métricas miden cosas distintas, una mide la precisión de campo y la otra mide el reparto membrana-flexión.

Esto explica el resultado. PB-PUNet acierta mejor las magnitudes absolutas de los campos, pero distribuye su error de una forma que distorsiona el cociente. EquiNO+σ^2, al concentrar su capacidad en la estructura de baja frecuencia que determina ese reparto, predice mejor el balance membrana-flexión, que es justo lo que necesita el guiado, aunque pague un poco de precisión de campo absoluto.

2.2. El mismo patrón en geometrías con hueco

Para comprobar que la ventaja del operador espectral no depende del régimen, se repite el benchmark sobre geometrías con hueco, evaluando solo sobre el material con métricas enmascaradas. La Figura 5.3 muestra el mismo patrón, ya que EquiNO+σ^2 vuelve a tener el menor error de factor de membrana en todo el espectro de ruido, mientras que el baseline convolucional conserva su ventaja en el error de campo crudo a ruido bajo.

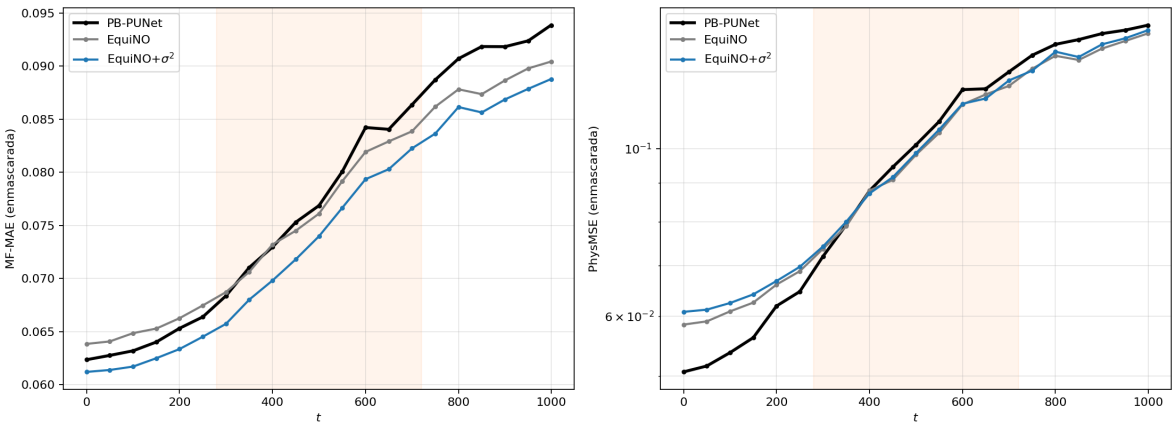

Figura 5.3: Benchmark enmascarado en geometrías con hueco (solo material). Izquierda: EquiNO+σ^2 tiene el menor error de factor de membrana en toda la ventana de ruido. Derecha: en el error de campo crudo, el baseline convolucional mantiene ventaja a ruido bajo. El reparto por métrica es el mismo que en *solid*.

La conclusión es robusta a través de regímenes, ya que para el factor de membrana, que es la magnitud que guía, el surrogado espectral con incertidumbre es la mejor opción tanto en *solid* como en *hole*.

3. Por qué funciona: el régimen de información de la difusión

¿Por qué un operador espectral ligero bate a una U-Net convolucional mucho mayor? La respuesta está en qué información sobrevive al ruido. Para cada timestep se mide la relación señal-ruido por modo de Fourier de la geometría. La Figura 5.4 muestra que la SNR colapsa de las altas a las bajas frecuencias al crecer el ruido, de forma que en estado limpio sobreviven modos hasta $k^\star \approx 22$, pero en plena ventana de guiado solo quedan por encima del umbral los uno o dos modos más bajos.

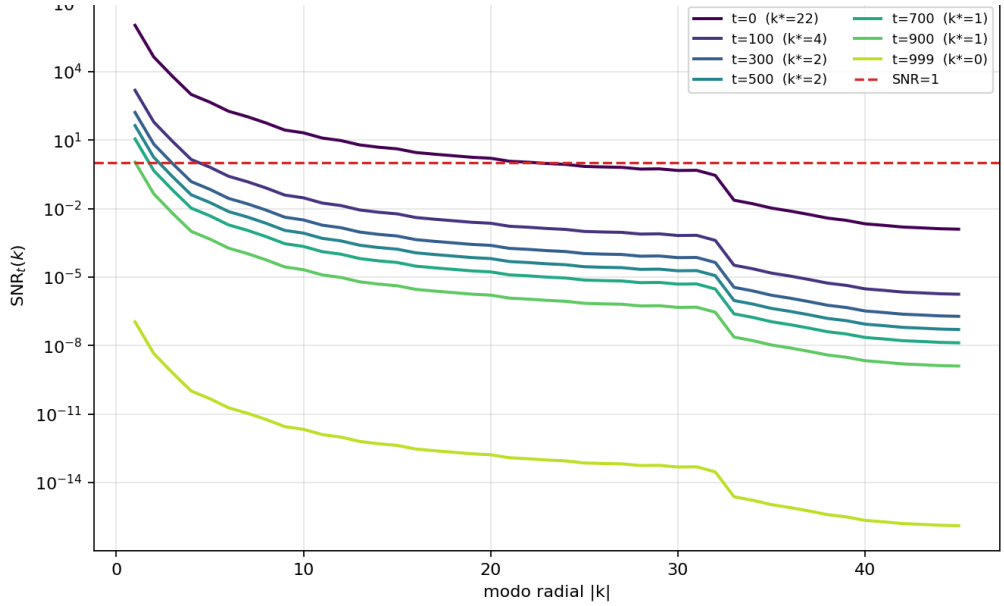

Figura 5.4: Relación señal-ruido por modo radial $|k|$ para distintos timesteps. Al crecer t, el último modo con SNR mayor que uno (marcado $k^\star$) cae de 22 a 1. En la ventana de guiado solo la información de baja frecuencia es recuperable desde el estado ruidoso.

Esto explica la eficiencia de EquiNO en la ventana de guiado. Al centrarse en los modos de baja frecuencia, aprovecha la información global que todavía permanece reconocible y evita dedicar capacidad a detalles de alta frecuencia que el ruido ya ha ocultado. PB-PUNet, en cambio, dispone de una arquitectura especialmente preparada para representar esos detalles locales, aunque en esta etapa apenas estén presentes. Por ello, una parte importante del error restante está determinada por la información perdida durante la difusión y no puede corregirse simplemente aumentando el tamaño del surrogado. En consecuencia, resulta más útil aprovechar mejor la señal disponible que recurrir a modelos de mayor capacidad.

4. La brecha de Jensen, demostrada empíricamente

La física que interesa, el factor de membrana, está definida sobre la geometría limpia, pero el muestreo opera sobre estados ruidosos. La pregunta es dónde evaluar la respuesta mecánica. Una opción habitual es estimar la geometría limpia con la fórmula de Tweedie, $\hat{x}_0 \approx \mathbb{E}[x_0 \mid x_t]$, y evaluar un surrogado limpio sobre esa estimación. El problema es que la física es no lineal, y evaluar el operador sobre la media posterior no equivale a la respuesta media posterior, y esa diferencia es la brecha de

Jensen. La alternativa que adopta este trabajo es entrenar un Ingeniero condicionado al tiempo que aprende directamente la respuesta media compatible con el estado ruidoso, $\mathbb{E}[R(x_0) \mid x_t, t]$, evitando así el paso intermedio que introduce el sesgo.

La Figura 5.5 compara empíricamente las tres opciones a lo largo de la trayectoria. El experimento se realiza con el surrogado convolucional de referencia (PB-PUNet) en sus tres variantes de evaluación, de modo que la comparación aísla la **estrategia** de evaluación y no la arquitectura, y el error se mide en unidades físicas crudas, sin normalizar. El panel superior mide la precisión del surrogado frente al ruido. El Ingeniero condicionado al tiempo (azul) y el surrogado limpio evaluado sobre la estimación de Tweedie (amarillo) son mucho más precisos que evaluar el surrogado limpio directamente sobre el estado ruidoso (rojo, evidentemente inviable). Pero, sobre todo, en la zona de ruido alto el Ingeniero condicionado al tiempo iguala o mejora a la ruta de Tweedie, precisamente porque no arrastra la brecha de Jensen. El panel inferior izquierdo lo confirma en la práctica del guiado, ya que la trayectoria guiada con el Ingeniero condicionado al tiempo alcanza antes un factor de membrana alto y lo mantiene, mientras que las otras dos rutas tardan más o se quedan por debajo.

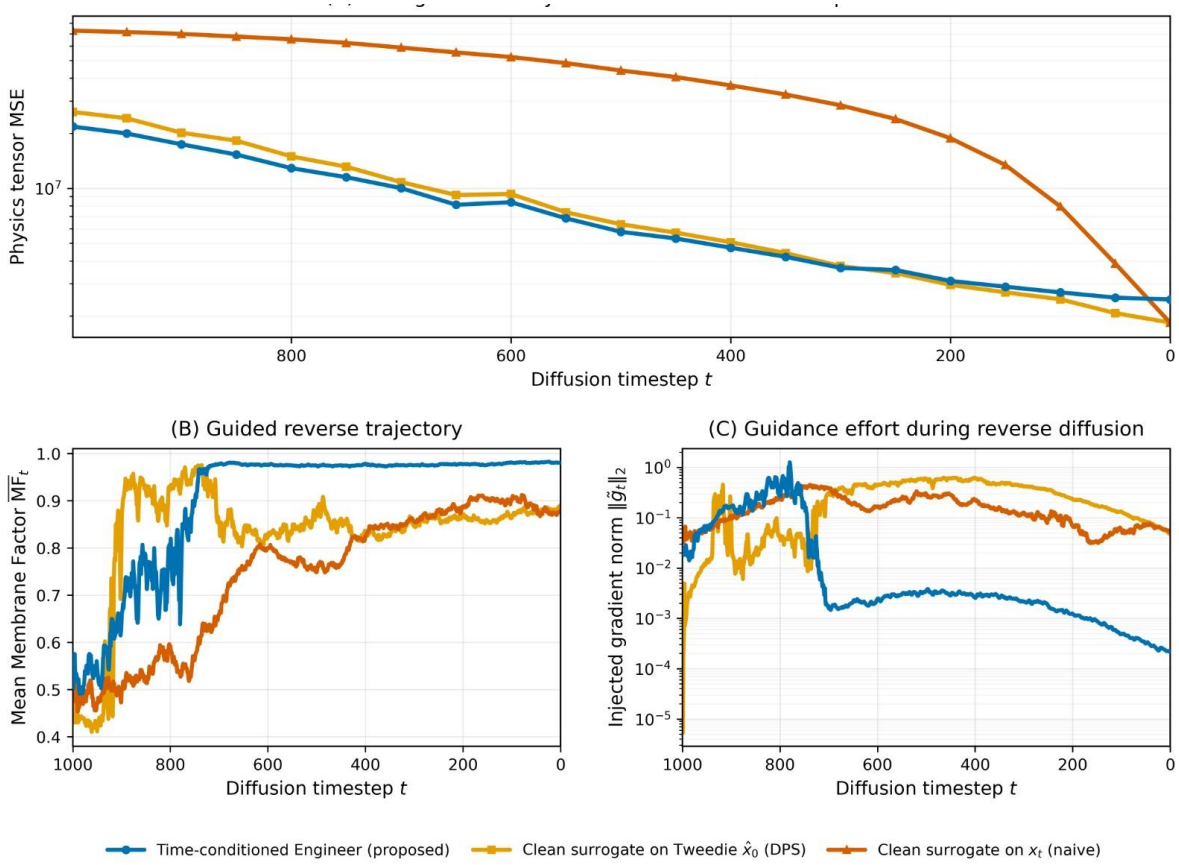

Figura 5.5: Demostración empírica del efecto de la brecha de Jensen, realizada con el surrogado convolucional de referencia (PB-PUNet) y error en unidades físicas crudas. Arriba: precisión del surrogado a lo largo del ruido para tres estrategias de evaluación: el Ingeniero condicionado al tiempo, el surrogado limpio evaluado sobre la estimación de Tweedie $\hat{x}_0$ y el surrogado limpio evaluado de forma ingenua sobre x_t. Abajo: factor de membrana medio a lo largo de la trayectoria guiada (B) y esfuerzo de gradiente inyectado (C). Aprender la respuesta media condicionada al estado ruidoso evita el sesgo de evaluar la física sobre la media posterior.

Al entrenar el surrogado para predecir la respuesta media condicionada a x_t, se elimina el error de primer orden. La diferencia restante depende de la no linealidad del objetivo y de la varianza asociada al estado ruidoso, por lo que aumenta a medida que crece el nivel de ruido. Este error residual puede estimarse combinando la curvatura del objetivo con la varianza predicha por la cabeza de incertidumbre.

5. Una incertidumbre calibrada y utilizable

Para que la incertidumbre del surrogado sirva como señal de guiado tiene que cumplir dos cosas, crecer con el ruido y ordenar el error. La Figura 5.6 muestra que ambas se cumplen. En el panel izquierdo, la varianza predicha aumenta de forma monótona con el timestep, en paralelo al error real, aunque por debajo de él, ya que la red ordena bien el error pero infraestima su escala. Esa infraestimación se mitiga con una recalibración por temperatura, $\tilde{\sigma}^2 = \tau\,\sigma_\phi^2$. La temperatura $\tau = 1{,}78$ se obtiene de forma empírica, ajustándola en una mitad del conjunto de validación para igualar la escala de la varianza al error medido, y evaluándola en la otra mitad. La mejora tiene un porqué y un límite claros. Como τ es un único factor global, corrige el nivel medio de la varianza (y por eso el diagrama de fiabilidad del panel derecho se aproxima a la diagonal), pero no puede corregir su **forma** temporal. Se aprecia en la propia figura, ya que la curva recalibrada queda cerca del error en la zona intermedia, todavía por debajo en los estados casi limpios y por encima en el extremo de ruido alto. Para el uso que se le da aquí es suficiente, porque el schedule de guiado consume el perfil relativo de la incertidumbre, no su valor absoluto. Aun así, es una calibración deliberadamente simple, y así se recoge en las limitaciones y el trabajo futuro.

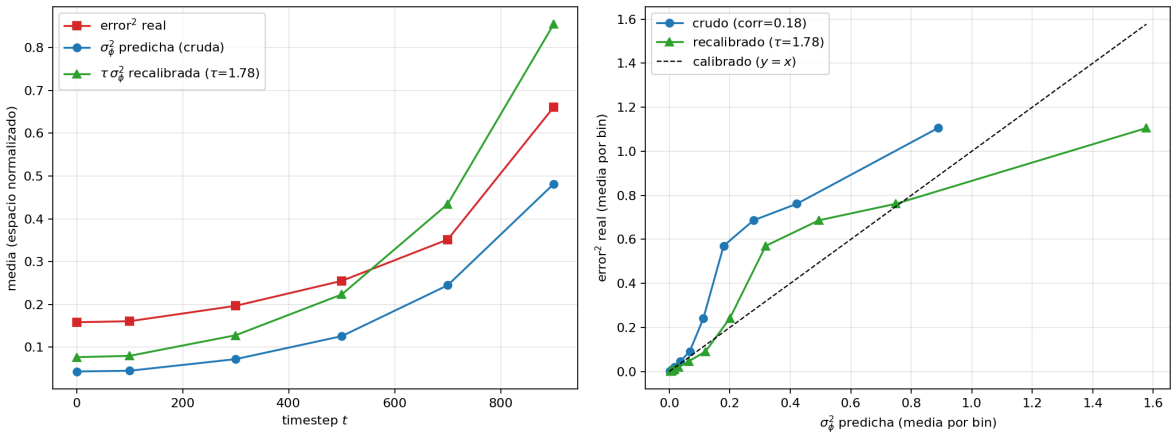

Figura 5.6: Calibración de la incertidumbre de EquiNO+σ^2. Izquierda: la varianza predicha cruda (azul) sigue al error real (rojo) pero lo infraestima; la varianza recalibrada con $\tau = 1{,}78$ (verde) se acerca al error. Derecha: diagrama de fiabilidad, donde la recalibración aproxima los puntos a la diagonal $y = x$.

Conviene distinguir dos niveles de la señal. Píxel a píxel, la correlación entre la varianza y el error es modesta (en torno a 0,18). Pero el guiado adaptativo no usa la varianza píxel a píxel, sino su perfil

promedio por timestep, y a ese nivel, que es el que de verdad se utiliza, la señal es muy alta, ya que la correlación entre el perfil u_t y el error medio por timestep es de 0,997, evaluada sobre la rejilla de timesteps de calibración (seis niveles de ruido, por lo que el valor debe leerse como consistencia de dos curvas monótonas más que como una correlación fina). Es decir, la incertidumbre agregada reproduce con enorme fidelidad cómo crece el error con el ruido, que es exactamente lo que el schedule necesita.

6. Generación guiada con el modelo propuesto

Una vez definidos los componentes del método, se generan las láminas a partir del prior congelado y se aplica el guiado mediante EquiNO+σ^2. Se comparan, con semillas pareadas, tres condiciones, la generación sin guiado, el guiado con una campana temporal fija y el guiado adaptativo según la incertidumbre. La calidad se evalúa mediante el factor de membrana predicho por EquiNO+σ^2, al ser el surrogado más preciso del estudio. No obstante, dado que este mismo modelo interviene tanto en el guiado como en la evaluación, la posible circularidad de esta elección se considera una limitación y se discute en el Capítulo 6.

La Figura 5.7 resume el resultado. El prior sin guiar genera láminas con un factor de membrana medio de 0,53, y el guiado lo eleva hasta 0,87. La fracción de láminas con factor de membrana superior a 0,8 pasa del 17 % sin guiar al 87 % con guiado, y la fracción por encima de 0,9 pasa del 3 % a casi la mitad.

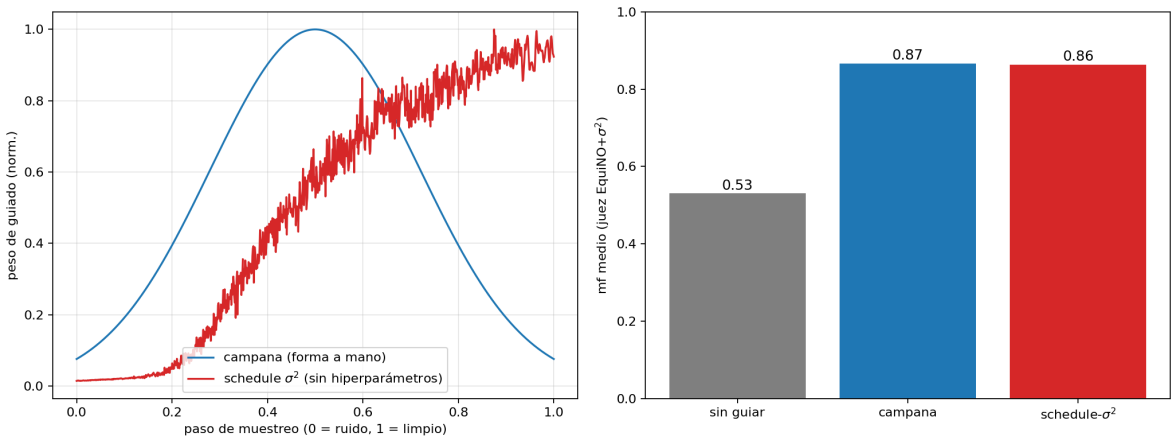

Figura 5.7: Generación guiada en *solid*. Izquierda: la campana fija impone una forma a mano que tiene su pico en mitad de la trayectoria, mientras que el schedule deduce su forma del perfil de incertidumbre y concentra el guiado cerca de los estados limpios, donde la red es más fiable. Derecha: factor de membrana medio de las láminas generadas, juzgado por EquiNO+σ^2. El guiado eleva claramente el factor de membrana frente al prior sin guiar.

Tabla 5.2: Generación guiada en *solid* ($N = 256$, muestreo a 1000 pasos), juzgada por EquiNO+σ^2. El *convex hull* mide la diversidad de las geometrías en el latente.

Estrategia	mf medio	$P(\text{mf} > 0{,}8)$	$P(\text{mf} > 0{,}9)$	Convex hull
Sin guiado	0.53	0.17	0.03	24.9
Campana fija	**0.87**	**0.87**	0.43	26.4
Schedule-σ^2	0.86	**0.87**	**0.48**	**29.1**

El resultado es doble, y las diferencias se cuantifican con bootstrap pareado sobre las 256 muestras (mismas semillas entre estrategias, $B = 5000$). Primero, el guiado funciona, ya que lleva el factor de membrana medio de 0,53 a 0,87 y casi todas las láminas superan el umbral de 0,8. Segundo, y es la aportación metodológica, el schedule derivado de la incertidumbre iguala estadísticamente a la campana ajustada a mano sin usar ningún hiperparámetro de forma, con una diferencia de factor de membrana de $-0{,}003$ con IC del 95 % $[-0{,}021, +0{,}015]$, un empate claro. Las señales de mayor diversidad apuntan a favor del schedule pero no alcanzan significancia individual, ya que el *convex hull* sube de 26,4 a 29,1 (diferencia +3,5, IC $[-1{,}3, +9{,}8]$) y la fracción de láminas por encima de 0,9 de 0,43 a 0,48 (diferencia +0,06, IC $[-0{,}03, +0{,}14]$). Los resultados no demuestran que el *schedule* adaptativo supere a una campana temporal bien ajustada, pero sí muestran que alcanza una mejora comparable sin necesidad de calibración manual. Además, presenta señales claras de favorecer una mayor diversidad en las geometrías generadas. De este modo, la incertidumbre permite regular

automáticamente el equilibrio entre la fiabilidad del guiado y la libertad de exploración durante la generación. Esa cobertura se aprecia en la Figura 5.8, que dibuja cada estrategia en el espacio latente de diseño.

Como control de circularidad, la comparación se repite utilizando PB-PUNet como evaluador independiente. Los resultados mantienen la misma tendencia, ya que el factor de membrana aumenta de 0,63 sin guiado a 0,90 con campana fija y 0,87 con el *schedule* adaptativo. Por tanto, la mejora no depende exclusivamente de EquiNO+σ^2, que es el modelo empleado durante el guiado.

No obstante, estas cifras deben interpretarse con cautela. PB-PUNet tiende a sobrestimar el valor absoluto del factor de membrana en las geometrías generadas, en línea con el sesgo observado sobre el conjunto de validación, +0,045, frente a +0,016 de EquiNO+σ^2. Por ello, se utiliza principalmente para confirmar la dirección de la mejora y no como referencia de su magnitud exacta. Además, bajo este evaluador, la campana ajustada mantiene una ventaja pequeña pero significativa sobre el *schedule* adaptativo, con una diferencia de $-0{,}029$ e intervalo de confianza $[-0{,}047, -0{,}012]$. Este resultado refuerza la idea de que el *schedule* obtiene una mejora próxima a la de una campana afinada, aunque cede ligeramente en calidad a cambio de eliminar el ajuste manual.

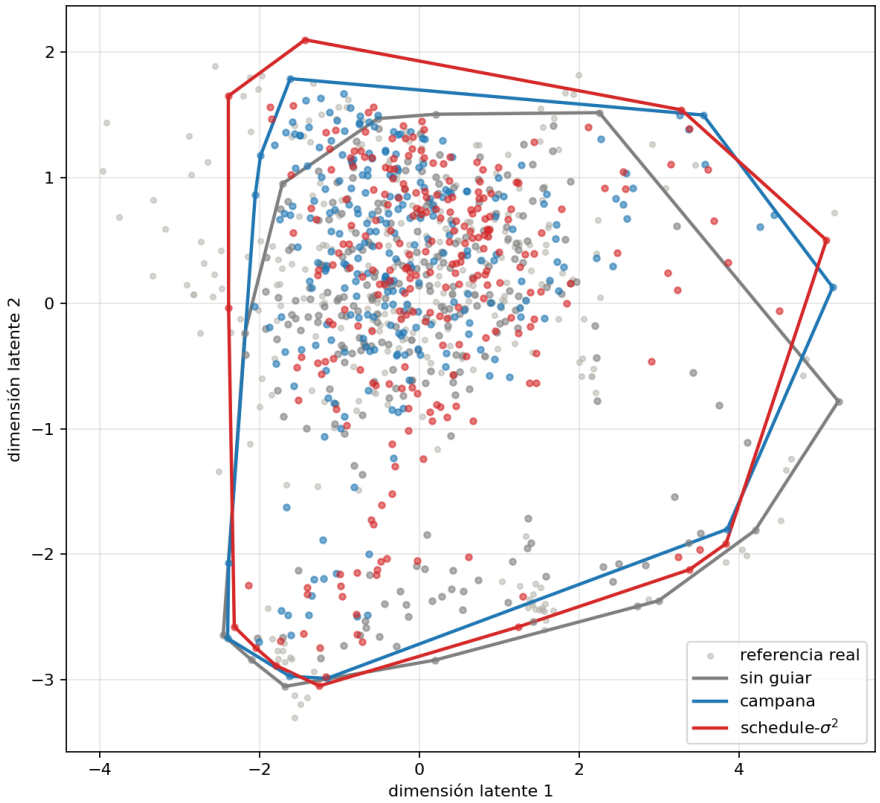

Figura 5.8: Cobertura del espacio de diseño en *solid*: *convex hull* de las geometrías generadas sobre una proyección 2D del latente VAE. Cada nube son las láminas de una estrategia y el polígono su envolvente convexa; en gris, la referencia de láminas reales. El schedule-σ^2 (rojo) cubre algo más de área que la campana, es decir, explora diseños más diversos.

El efecto del guiado se ve de un vistazo en la Figura 5.9, que compara láminas del prior sin guiar con láminas guiadas por EquiNO+σ^2 con el schedule de incertidumbre, coloreadas por su factor de membrana local. En las láminas sin guiar abundan las regiones de flexión (rojo y naranja) y su factor de membrana medio es bajo, mientras que en las guiadas predomina el trabajo de membrana (verde) y las concentraciones de flexión quedan confinadas a apoyos localizados, con un factor de membrana muy superior. Las formas guiadas siguen siendo diversas entre sí, de modo que la mejora mecánica no colapsa el repertorio de geometrías.

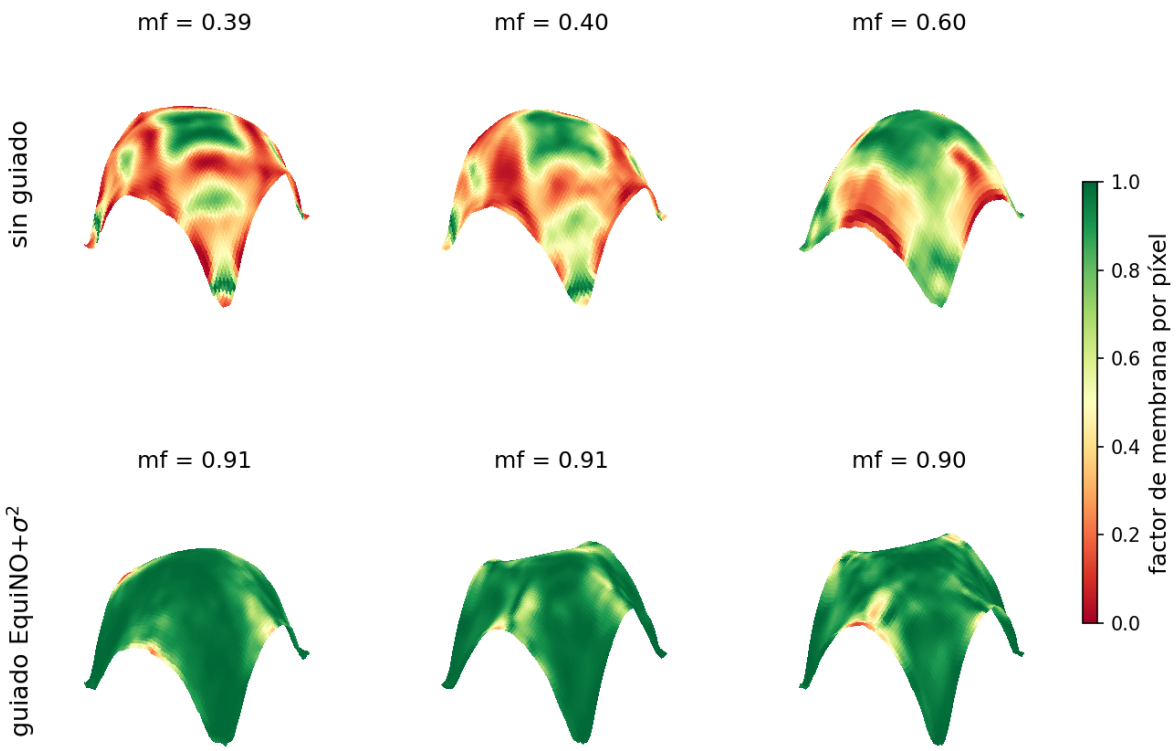

Figura 5.9: Láminas generadas en *solid*, coloreadas por el factor de membrana local (verde: membrana; rojo: flexión). Arriba, tres muestras del prior sin guiar; abajo, tres muestras guiadas por EquiNO+σ^2 con el schedule de incertidumbre. El guiado transforma superficies con amplias zonas de flexión (mf $\approx$ 0,4–0,6) en láminas casi enteramente de membrana (mf $\approx$ 0,9), manteniendo formas variadas.

6.1. Fuerza de guiado y compromiso calidad-diversidad

Hasta aquí se ha estudiado la **forma** temporal del guiado, y queda su **intensidad**. El parámetro γ que escala el gradiente controla cuánta corrección física se inyecta, y barrerlo a lo largo de más de un orden de magnitud revela un compromiso suave entre calidad y diversidad (Figura 5.10 y Tabla 5.3). El barrido usa exactamente el mismo protocolo que el resto del capítulo ($N = 256$, muestreo fiel de 1000 pasos, semillas pareadas, campana fija), de modo que sus extremos coinciden con estrategias ya presentadas, ya que $\gamma = 0$ es el prior sin guiar y $\gamma = 600$ es la campana de referencia de la Tabla 5.2. Sin guiado, el factor de membrana medio es 0,53, y al aumentar γ crece de forma monótona hasta 0,93 con $\gamma = 2000$, y la fracción de láminas por encima de 0,9 pasa del 3 % al 81 %, con intervalos de confianza estrechos en toda la curva.

La diversidad, medida por el convex hull, dibuja en cambio una curva no monótona y con más

variabilidad. Un guiado suave ($\gamma = 100$) apunta a una diversidad incluso mayor que la del prior sin guiar (30,2 frente a 24,9, aunque los intervalos se solapan parcialmente), coherente con la idea de que una corrección moderada empuja a las muestras fuera de la moda del prior hacia formas variadas de factor de membrana alto. A partir de ahí la tendencia es descendente, con fluctuaciones propias de la métrica, y con $\gamma = 2000$ la cobertura cae a 17,9, ya que el guiado fuerte concentra la generación en un repertorio reducido de geometrías muy eficientes.

Los intervalos asociados al área de la envolvente convexa requieren una interpretación específica. Como una submuestra no puede cubrir una región mayor que la muestra completa salvo por variaciones particulares del muestreo, el *bootstrap* tiende a infraestimar esta métrica. Por ello, el valor calculado con las N muestras suele situarse cerca del límite superior del intervalo e incluso puede quedar por encima de él, como ocurre en varias filas de la Tabla 5.3. Estos intervalos se utilizan, por tanto, para describir la variabilidad del área estimada y no como intervalos centrados alrededor del valor observado. Las comparaciones estadísticas entre estrategias se realizan mediante diferencias pareadas, de forma que este sesgo afecta de manera similar a ambas condiciones y se reduce en primer orden.

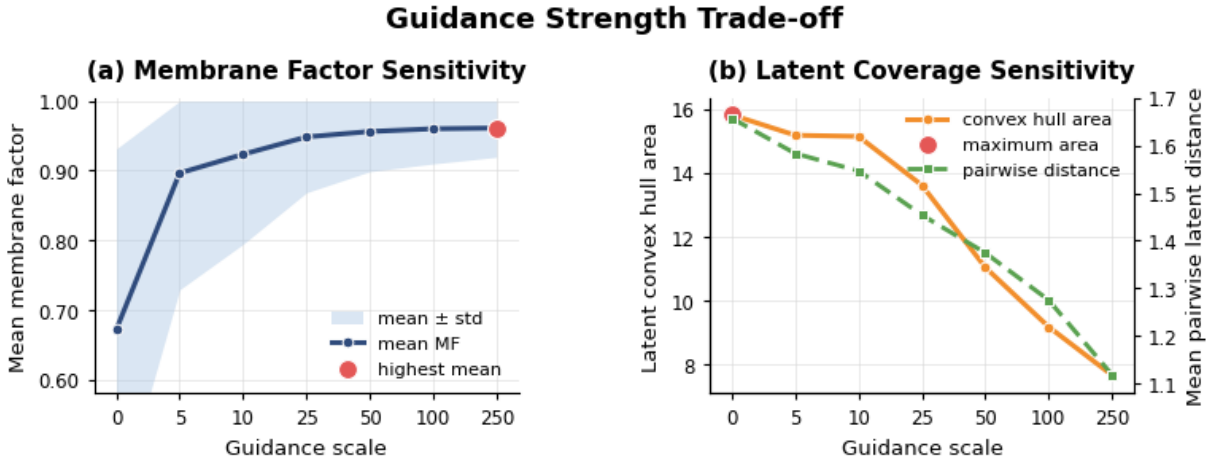

Figura 5.10: Compromiso calidad–diversidad al variar la fuerza de guiado en un barrido complementario. (a) El factor de membrana medio crece de forma monótona con la escala de guiado y satura (banda: media ± desviación típica; en rojo, el punto de mayor media). (b) La cobertura del espacio latente, medida por el área de *convex hull* (naranja) y por la distancia media par a par (verde), decae al aumentar el guiado. Las dos métricas de diversidad son consistentes entre sí. Los ejes están rotulados en inglés (*guidance scale*, escala de guiado; *mean membrane factor*, factor de membrana medio; *latent convex hull area*, área de *convex hull* latente; *mean pairwise latent distance*, distancia latente media par a par).

Tabla 5.3: Barrido de la fuerza de guiado en *solid* (EquiNO+σ^2, campana fija, $N = 256$, 1000 pasos, semillas pareadas). Entre corchetes, intervalos de confianza al 95 % por bootstrap. Las filas $\gamma = 0$ y $\gamma = 600$ son los mismos brazos que la Tabla 5.2.

γ	mf medio [IC 95 %]	$P(\text{mf} > 0{,}8)$	$P(\text{mf} > 0{,}9)$	convex hull [IC 95 %]
0	0.53 [0.51, 0.56]	0.17	0.03	24.9 [21.0, 25.2]
100	0.76 [0.74, 0.78]	0.56	0.13	**30.2** [22.5, 30.4]
250	0.83 [0.81, 0.84]	0.77	0.24	24.7 [19.7, 24.7]
500	0.86 [0.85, 0.87]	0.86	0.34	18.4 [14.3, 18.8]
600	0.87 [0.86, 0.88]	0.87	0.43	26.4 [18.0, 26.5]
1000	0.89 [0.88, 0.90]	0.93	0.57	21.8 [16.0, 22.0]
2000	**0.93** [0.92, 0.94]	0.98	0.81	17.9 [11.5, 23.1]

El resultado es, por tanto, una interpolación continua entre dos regímenes, sin ningún cambio de arquitectura ni de entrenamiento, ya que con γ moderado el sistema preserva, y probablemente amplía, la diversidad a cambio de una calidad intermedia, y con γ alto prioriza la calidad sacrificando exploración. El diseñador puede así elegir el punto de operación según necesite explorar alternativas o maximizar la eficiencia, simplemente ajustando un escalar en la inferencia.

Este mismo compromiso se aprecia de forma independiente en un barrido complementario (Figura 5.10), que separa las dos caras del fenómeno. El panel de sensibilidad del factor de membrana muestra la calidad creciendo de forma monótona con la escala de guiado y saturando, mientras que el panel de cobertura latente muestra la diversidad, medida a la vez por el área de *convex hull* y por la distancia media par a par en el espacio latente, decayendo al aumentar el guiado. La coincidencia entre ambas métricas de diversidad refuerza la conclusión de que un guiado más intenso concentra la generación y reduce la variedad de las geometrías obtenidas. Los valores del punto de operación del Ingeniero estudiado, con sus intervalos, son los de la Tabla 5.3.

6.2. Generación en geometrías con hueco

La generación con hueco requiere dos consideraciones adicionales. El factor de membrana se calcula únicamente sobre la región ocupada por material, mientras que el guiado se restringe mediante la geometría limpia estimada en cada paso. Así se evita que el modelo mejore artificialmente la métrica rellenando el agujero. Los resultados confirman la eficacia de esta medida, ya que la fracción de hueco se mantiene alrededor del 8 % en todas las estrategias y el guiado conserva la abertura durante la generación.

El prior sin guiar ya genera láminas cuyo material es mayoritariamente de membrana, con un factor enmascarado de 0,76, en línea con la referencia de datos reales ($\approx 0{,}75$). El margen de mejora es por tanto pequeño. Aun así, el guiado con EquiNO+σ^2 eleva el factor enmascarado a 0,80, y el schedule-σ^2 vuelve a igualar a la campana (0,79) con mayor diversidad, reproduciendo la misma historia que en *solid* (Tabla 5.4 y Figura 5.12). La Figura 5.11 muestra ejemplos generados, y el guiado respeta el hueco (no lo rellena) y mejora el factor de membrana enmascarado, aunque partiendo ya de un valor alto el salto visual es más sutil que en *solid*.

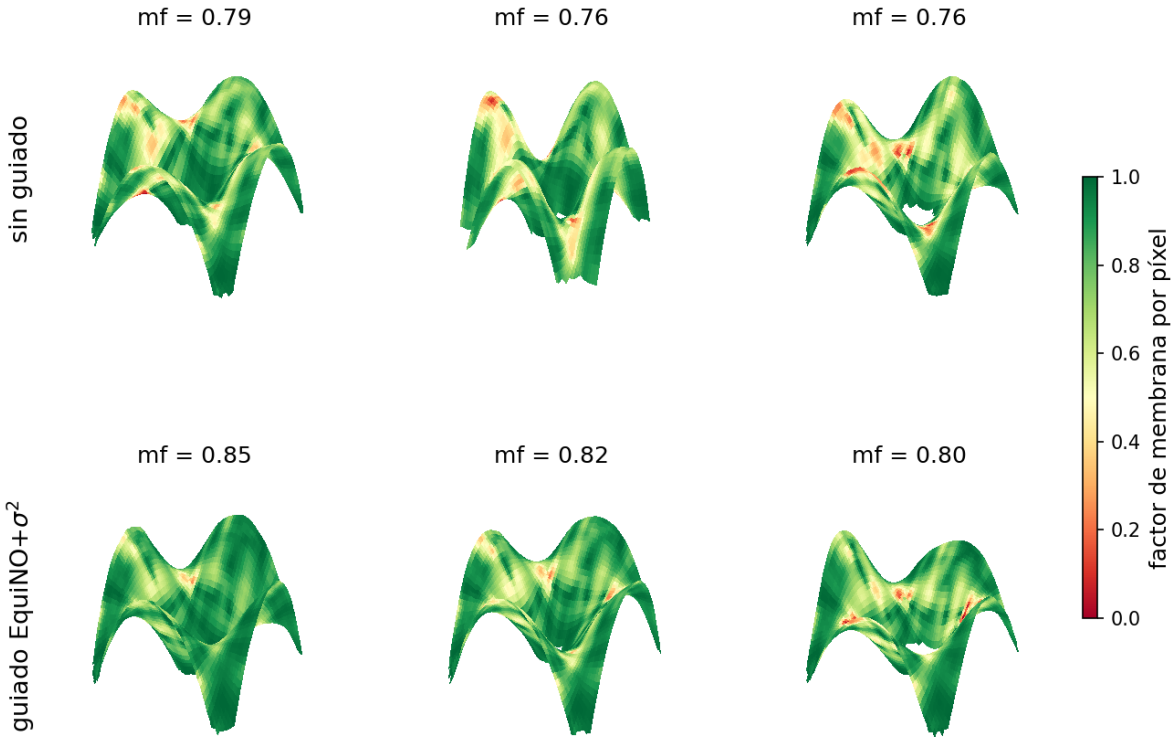

Figura 5.11: Láminas generadas en *hole*, coloreadas por el factor de membrana local enmascarado (la región sin material se deja hueca). Arriba, prior sin guiar; abajo, guiado por EquiNO+σ^2 con el schedule de incertidumbre. El guiado conserva el hueco y sube el factor de membrana del material; el margen es menor que en *solid* porque el prior ya parte de láminas muy funiculares.

Tabla 5.4: Generación guiada en *hole* ($N = 256$, 1000 pasos), con guiado y métrica enmascarados; juez: EquiNO+σ^2, como en todo el capítulo.

Estrategia	mf enmascarado (EquiNO+σ^2)	fracción de hueco	convex hull
Sin guiado	0.76	8.4 %	16.7
Campana fija	**0.80**	8.0 %	16.0
Schedule-σ^2	0.79	8.5 %	**17.6**

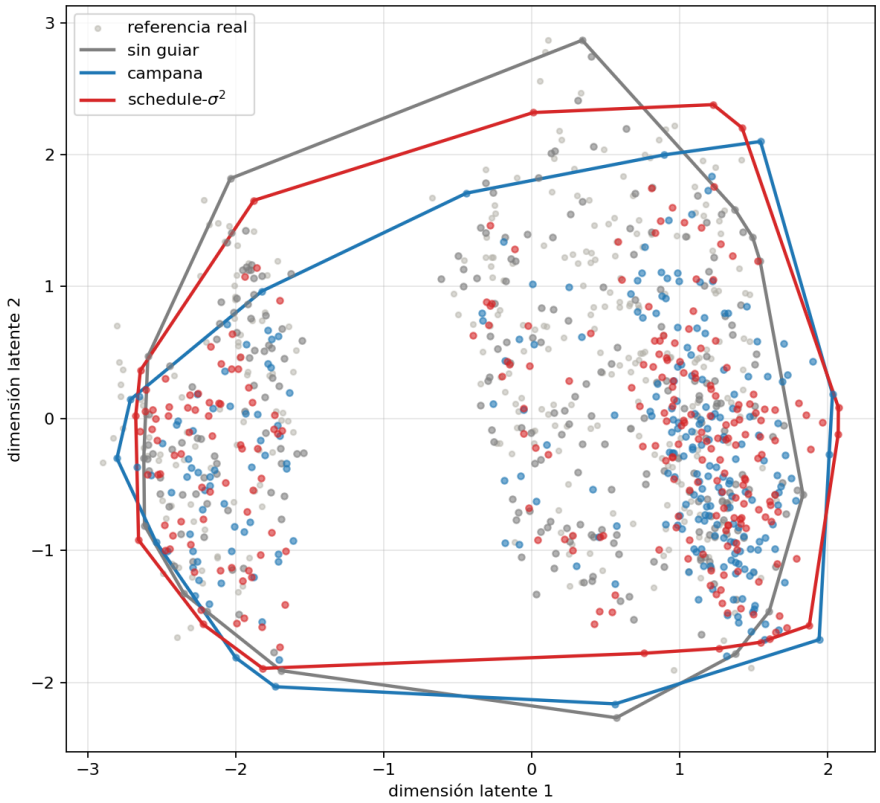

Figura 5.12: Cobertura del espacio de diseño en *hole*: *convex hull* de las geometrías generadas sobre el latente de un VAE entrenado en geometrías con hueco. Como en *solid*, el schedule-σ^2 (rojo) explora algo más que la campana.

En el régimen *hole*, el evaluador independiente PB-PUNet también confirma la mejora observada con EquiNO+σ^2. Mientras el modelo espectral estima un aumento del factor de membrana tras aplicar el guiado, PB-PUNet registra igualmente una mejora, desde 0,80 sin guiado hasta 0,90 con campana. Esta coincidencia, al igual que en el régimen *solid*, indica que el incremento no depende exclusivamente del modelo utilizado para guiar y evaluar. Por tanto, el control con una arquitectura de familia distinta reduce la posible circularidad y refuerza la conclusión de que el guiado mejora la calidad mecánica de las geometrías generadas con hueco. Además, los resultados de precisión presentados en la Sección 2 muestran que EquiNO+σ^2 obtiene el menor error en la estimación del factor de membrana enmascarado.

7. Discusión

Los resultados presentan una historia consistente. La ventaja de EquiNO+σ^2 sobre el baseline convolucional se mantiene a lo largo de todos los experimentos, y la incertidumbre predicha por el Ingeniero condicionado al tiempo permite un guiado adaptativo que iguala a una campana ajustada a mano sin necesidad de calibración manual. El guiado aunque no mejora claramente en factor de membrana a la campana, sí apunta a una mayor diversidad de geomentrías, además de conseguir un equilibrio automático entre calidad y diversidad al variar la fuerza de guiado. La ventaja del surrogado está presente incluso en geometrías con hueco, lo que confirma que la superioridad del operador espectral no depende del régimen. La brecha de Jensen se demuestra empíricamente y se evita entrenando el surrogado para predecir la respuesta media condicionada al estado ruidoso, y la incertidumbre predicha es calibrable y útil para guiar la generación. El alcance de estos resultados se recogen de forma unificada en el Capítulo 6, donde se señalan limitaciones y líneas de trabajo futuro.

Capítulo 6

Conclusión y trabajo futuro

1. Conclusiones

Este trabajo fin de máster se articula en torno a dos ideas principales. La primera plantea combinar la capacidad de los modelos de difusión para generar geometrías diversas con la evaluación rápida y diferenciable que proporcionan los modelos surrogados, de forma que la generación pueda orientarse hacia soluciones mecánicamente más eficientes. La segunda estudia cómo aprovechar la incertidumbre inherente al proceso de difusión para adaptar el guiado durante la generación y decidir en qué momentos conviene intervenir con mayor o menor intensidad.

Sobre estas ideas se construye un pipeline modular, con un generador de difusión (el Arquitecto) y un surrogado diferenciable (el Ingeniero) entrenados por separado y acoplados solo durante la inferencia mediante un guiado diferenciable. La separación de responsabilidades evita la optimización conjunta, el equilibrio adversario y los reentrenamientos específicos de la tarea, y permite entrenar, validar y sustituir cada componente de manera independiente.

Además se ha demostrado matemáticamente y empíricamente que el surrogado aprende la respuesta mecánica esperada condicionada a cada estado ruidoso, abordando la brecha de Jensen que aparece al evaluar una física no lineal sobre una estimación limpia. La memoria formaliza esa brecha y la enmarca, junto con la relación señal-ruido espectral, en un análisis que explica por qué se producen los resultados observados.

Para atribuir las diferencias observadas únicamente a los componentes analizados, el Arquitecto se mantiene congelado durante todo el estudio. El trabajo se centra así en el Ingeniero y en el mecanismo de guiado, de los que se extraen tres conclusiones principales.

En primer lugar, EquiNO+σ^2, un operador espectral con estimación de incertidumbre, predice el factor de membrana con mayor precisión que el baseline convolucional en la ventana de ruido donde se aplica el guiado. Además, alcanza este resultado con aproximadamente 3,5 veces menos parámetros, lo que demuestra que una arquitectura más ligera puede aprovechar mejor la información disponible en los estados intermedios de la difusión.

En segundo lugar, esta ventaja se explica por la pérdida progresiva de información durante la trayectoria generativa. A medida que aumenta el ruido, las componentes de alta frecuencia desaparecen y la señal recuperable se concentra en los modos bajos. Por ello, el error del surrogado en la ventana de guiado depende principalmente de la incertidumbre introducida por el propio ruido y no de una falta de capacidad del modelo. El límite observado responde, por tanto, a la información que permanece disponible en cada estado y no a una arquitectura insuficiente.

En tercer lugar, la incertidumbre estimada por el surrogado resulta informativa y puede calibrarse. A partir de ella se construye un *schedule* temporal de guiado que alcanza una mejora comparable a la de una campana ajustada manualmente, sin necesidad de definir hiperparámetros para controlar su forma y manteniendo una diversidad ligeramente superior en las geometrías generadas.

Los experimentos de generación confirman además que el pipeline cumple su objetivo. La incorporación del guiado físico incrementa de forma clara el factor de membrana respecto al prior sin guiar. Al mismo tiempo, la fuerza de guiado permite recorrer de manera continua el compromiso entre calidad y diversidad. Los valores bajos preservan, e incluso pueden ampliar, la variedad del prior, mientras que los valores elevados concentran la generación en geometrías mecánicamente más eficientes. Este cambio se consigue sin modificar la arquitectura ni reentrenar las redes, por lo que el diseñador dispone de un control interpretable sobre el comportamiento del sistema según priorice la exploración de alternativas o la maximización del rendimiento mecánico.

En conjunto, los resultados muestran que la separación modular entre un prior geométrico y un surrogado consciente del ruido constituye una estrategia estable y eficaz para incorporar informa-

ción física en un proceso generativo. Esta arquitectura permite mejorar de forma independiente sus componentes principales mediante un surrogado más eficiente y preciso en la métrica utilizada para guiar y mediante una regla temporal que adapta automáticamente la intervención a partir de la incertidumbre estimada por el propio modelo.

2. Limitaciones

Por otro lado, conviene reconocer que el trabajo presenta limitaciones. La principal es que la evaluación de la generación no incorpora una simulación de elementos finitos nueva sobre las láminas generadas, por lo que se trata de una comparación relativa y no constituye una validación estructural definitiva. Para reducir la circularidad asociada al uso del mismo surrogado durante el guiado y la evaluación, se ha hecho uso de PB-PUNet como juez independiente, cuyo veredicto confirma la mejora del guiado en ambos regímenes.

Cada arquitectura se entrena con una única semilla, por lo que las diferencias observadas deben interpretarse en relación con esos entrenamientos concretos y no como una medida completa de la variabilidad entre ejecuciones.

La calibración de la incertidumbre se realiza mediante una única temperatura global, que corrige su escala media pero no su evolución a lo largo del proceso de difusión. Como resultado, la varianza recalibrada todavía tiende a infraestimar el error en los estados casi limpios y a sobreestimarlo en los niveles de ruido más altos. Esta aproximación resulta suficiente para construir el *schedule*, ya que este utiliza el perfil relativo de la incertidumbre, pero no permite interpretar $\tilde{\sigma}^2$ como una estimación probabilística perfectamente calibrada en cualquier *timestep*.

Finalmente, la ventaja del surrogado depende de la métrica considerada. EquiNO+σ^2 obtiene mejores resultados en el factor de membrana, que es la magnitud utilizada durante el guiado, pero no supera siempre al baseline en la reconstrucción de los campos mecánicos completos. Del mismo modo, el guiado adaptativo no alcanza una calidad superior a la de una campana temporal bien ajustada, sino una mejora comparable sin necesidad de definir manualmente sus parámetros de forma y con una diversidad ligeramente mayor.

3. Trabajo futuro

Las líneas de trabajo futuro se derivan directamente de las limitaciones identificadas. La principal consiste en incorporar una validación mecánica independiente mediante elementos finitos sobre una muestra de las láminas generadas. Esto permitiría confirmar de forma absoluta la mejora del factor de membrana y eliminar la circularidad asociada al uso de un surrogado como evaluador. También convendría ampliar el análisis de diversidad con métricas de referencia, como el Vendi Score (Friedman y Dieng, 2022), que complementen la información proporcionada por el área de la envolvente convexa.

Otra línea de mejora se encuentra en la construcción del *schedule* adaptativo. En este trabajo, la intensidad temporal del guiado se obtiene directamente de la precisión estimada en cada timestep. Futuras formulaciones podrían incorporar también la curvatura del objetivo mecánico para ajustar con mayor precisión cuándo y con qué intensidad debe intervenir el guiado. La calibración de la incertidumbre podría igualmente sustituir la temperatura global por métodos capaces de corregir su evolución temporal, como una transformación afín dependiente del timestep.

La memoria utiliza únicamente el perfil temporal medio de la incertidumbre, aunque EquiNO+σ^2 produce un mapa espacial completo $\sigma^2(x, y)$ para cada estado. Aprovechar esta información permitiría modular localmente el gradiente mediante un factor de confianza espacial o identificar las regiones de la geometría en las que la evaluación mecánica resulta menos fiable. Esta última posibilidad también facilitaría presentar al diseñador información interpretable sobre la fiabilidad local de las predicciones.

Finalmente, la estructura modular del pipeline facilita su aplicación a nuevas tipologías estructurales, condiciones de contorno y objetivos mecánicos. También permite plantear su integración en una herramienta de diseño interactivo, en la que el surrogado proporcione de forma inmediata tanto una evaluación mecánica como una estimación de su fiabilidad mientras el diseñador explora diferentes alternativas geométricas.

Referencias

Aage, N., Andreassen, E., Lazarov, B. S., y Sigmund, O. (2017). Giga-voxel computational morphogenesis for structural design. *Nature*, *550*, 84–86. doi: 10.1038/nature23911

Adriaenssens, S., Block, P., Veenendaal, D., y Williams, C. (Eds.). (2014). *Shell structures for architecture: Form finding and optimization*. London: Routledge. doi: 10.4324/9781315849270

Arjovsky, M., Chintala, S., y Bottou, L. (2017). Wasserstein generative adversarial networks. En *Proceedings of the 34th international conference on machine learning* (Vol. 70, pp. 214–223).

Bastek, J.-H., Sun, W., y Kochmann, D. M. (2024). Physics-informed diffusion models. *arXiv preprint*.

Bendsøe, M. P., y Sigmund, O. (2003). *Topology optimization: Theory, methods, and applications*. Berlin: Springer.

Bhooshan, S., Bhooshan, V., Dell'Endice, A., Chu, J., Singer, P., Megens, J., ... Block, P. (2022). The striatus bridge. *Architecture, Structures and Construction*, *2*, 521–543.

Block, P., y Ochsendorf, J. (2007). Thrust network analysis: A new methodology for three-dimensional equilibrium. *Journal of the International Association for Shell and Spatial Structures*, *48*(3), 167–173.

Chinesta, F., Cueto, E., Abisset-Chavanne, E., Duval, J. L., y El Khaldi, F. (2020). Virtual, digital and hybrid twins: A new paradigm in data-based engineering and engineered data. *Archives of Computational Methods in Engineering*, *27*, 105–134. doi: 10.1007/s11831-018-9301-4

Chung, H., Kim, J., McCann, M. T., Klasky, M. L., y Ye, J. C. (2023). Diffusion posterior sampling for general noisy inverse problems. En *International conference on learning representations*.

Ciarlet, P. G. (2000). *Mathematical elasticity, volume iii: Theory of shells*. Amsterdam: North-Holland.

Cueto, E., y Chinesta, F. (2023). Thermodynamics of learning physical phenomena. *Archives of Computational Methods in Engineering*, *30*, 4653–4666.

Dhariwal, P., y Nichol, A. (2021). Diffusion models beat GANs on image synthesis. En *Advances in neural information processing systems* (Vol. 34, pp. 8780–8794).

Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association*, *106*(496), 1602–1614. doi: 10.1198/jasa.2011.tm11181

Faber, C. (1963). *Candela: The shell builder*. New York: Reinhold Publishing.

Friedman, D., y Dieng, A. B. (2022). The vendi score: A diversity evaluation metric for machine learning. *arXiv preprint*.

Gal, Y., y Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. En *Proceedings of the 33rd international conference on machine learning* (Vol. 48, pp. 1050–1059).

Giannone, G., y Ahmed, F. (2023). Diffusing the optimal topology: A generative optimization approach. *arXiv preprint*.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y.

(2014). Generative adversarial nets. En *Advances in neural information processing systems* (Vol. 27).

He, J., Spokoyny, D., Neubig, G., y Berg-Kirkpatrick, T. (2019). Lagging inference networks and posterior collapse in variational autoencoders. En *International conference on learning representations*.

Hernández, Q., Badías, A., González, D., Chinesta, F., y Cueto, E. (2021). Structure-preserving neural networks. *Journal of Computational Physics*, *426*, 109950. doi: 10.1016/j.jcp.2020.109950

Herron, E., Rade, J., Jignasu, A., Ganapathysubramanian, B., Balu, A., Sarkar, S., y Krishnamurthy, A. (2023). Latent diffusion models for structural component design. *arXiv preprint*.

Ho, J., Jain, A., y Abbeel, P. (2020). Denoising diffusion probabilistic models. En *Advances in neural information processing systems* (Vol. 33, pp. 6840–6851).

Ho, J., y Salimans, T. (2022). Classifier-free diffusion guidance. *arXiv preprint*.

Huang, J., Yang, G., Wang, Z., y Park, J. J. (2024). DiffusionPDE: Generative PDE-solving under partial observation. En *Advances in neural information processing systems*.

Huerta, S. (2006). Structural design in the work of gaud'i. *Architectural Science Review*, *49*(4), 324–339. doi: 10.3763/asre.2006.4943

Isler, H. (1961). New shapes for shells. *Bulletin of the International Association for Shell Structures*, *8*, 123–130. (Originally presented at the First Congress of the IASS, Madrid, 1959)

Jacobsen, C., Zhuang, Y., y Duraisamy, K. (2023). Cocogen: Physically-consistent and conditioned score-based generative models for forward and inverse problems. *arXiv preprint*.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620–630. doi: 10.1103/PhysRev.106.620

Kendall, A., y Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? En *Advances in neural information processing systems* (Vol. 30).

Kingma, D. P., y Welling, M. (2014). Auto-encoding variational bayes. En *International conference on learning representations*.

Kuleshov, V., Fenner, N., y Ermon, S. (2018). Accurate uncertainties for deep learning using calibrated regression. En *Proceedings of the 35th international conference on machine learning* (Vol. 80, pp. 2796–2804).

Lakshminarayanan, B., Pritzel, A., y Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. En *Advances in neural information processing systems* (Vol. 30).

Larsen, A. B. L., Sønderby, S. K., Larochelle, H., y Winther, O. (2016). Autoencoding beyond pixels using a learned similarity metric. En *Proceedings of the 33rd international conference on machine learning* (Vol. 48, pp. 1558–1566). PMLR.

Levi, D., Gispan, L., Giladi, N., y Fetaya, E. (2019). Evaluating and calibrating uncertainty prediction in regression tasks. *arXiv preprint*.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., y Anandkumar, A. (2021). Fourier neural operator for parametric partial differential equations. En *International conference on learning representations*.

Loshchilov, I., y Hutter, F. (2019). Decoupled weight decay regularization. En *International conference on learning representations*.

Lourenço, R., Alfaro, I., Moya, B., y Cueto, E. (2024). Physics-informed, generative adversarial design of funicular shells. *SSRN Electronic Journal*. doi: 10.2139/ssrn.6614732

Lu, L., Jin, P., Pang, G., Zhang, Z., y Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, *3*, 218–229. doi: 10.1038/s42256-021-00302-5

Malarz, D., Kasymov, A., Zięba, M., Tabor, J., y Spurek, P. (2025). Classifier-free guidance with adaptive scaling. *arXiv preprint*.

Mazé, F., y Ahmed, F. (2023). Diffusion models beat GANs on topology optimization. En *Proceedings of the aaai conference on artificial intelligence*.

Metz, L., Poole, B., Pfau, D., y Sohl-Dickstein, J. (2017). Unrolled generative adversarial networks. En *International conference on learning representations*.

Nichol, A., y Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. En *Proceedings of the 38th international conference on machine learning* (Vol. 139, pp. 8162–8171).

Papalampidi, P., Wiles, O., Ktena, I., Shtedritski, A., Bugliarello, E., Kajic, I., ... Nematzadeh, A. (2025). Dynamic classifier-free diffusion guidance via online feedback. *arXiv preprint*.

Reddy, J. N. (2007). *Theory and analysis of elastic plates and shells* (2.ª ed.). Boca Raton, FL: CRC Press.

Regenwetter, L., Nobari, A. H., y Ahmed, F. (2022). Deep generative models in engineering design: A review. *Journal of Mechanical Design*, *144*(7), 071704. doi: 10.1115/1.4053859

Regenwetter, L., Srivastava, A., Gutfreund, D., y Ahmed, F. (2023). Beyond statistical similarity: Rethinking metrics for deep generative models in engineering design. *arXiv preprint*.

Ronneberger, O., Fischer, P., y Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. En *Medical image computing and computer-assisted intervention (miccai)* (pp. 234–241). Springer.

Salimans, T., y Ho, J. (2022). Progressive distillation for fast sampling of diffusion models. En *International conference on learning representations*.

Sigmund, O., y Maute, K. (2013). Topology optimization approaches: A comparative review. *Structural and Multidisciplinary Optimization*, *48*, 1031–1055. doi: 10.1007/s00158-013-0978-6

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., y Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. En *Proceedings of the 32nd international conference on machine learning* (Vol. 37, pp. 2256–2265).

Song, J., Vahdat, A., Mardani, M., y Kautz, J. (2023). Pseudoinverse-guided diffusion models for inverse problems. En *International conference on learning representations*. Descargado de `https://jankautz.com/publications/PiGDM_ICLR23.pdf`

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., y Poole, B. (2021). Score-based generative modeling through stochastic differential equations. En *International conference on learning representations*.

Wangler, T., Roussel, N., Bos, F. P., Salet, T. A. M., y Flatt, R. J. (2019). Digital concrete: A review. *Cement and Concrete Research*, *123*, 105780. doi: 10.1016/j.cemconres.2019.105780

Zou, Z., Meng, X., y Karniadakis, G. E. (2023). Uncertainty quantification for noisy inputs-outputs in physics-informed neural networks and neural operators. *arXiv preprint*.