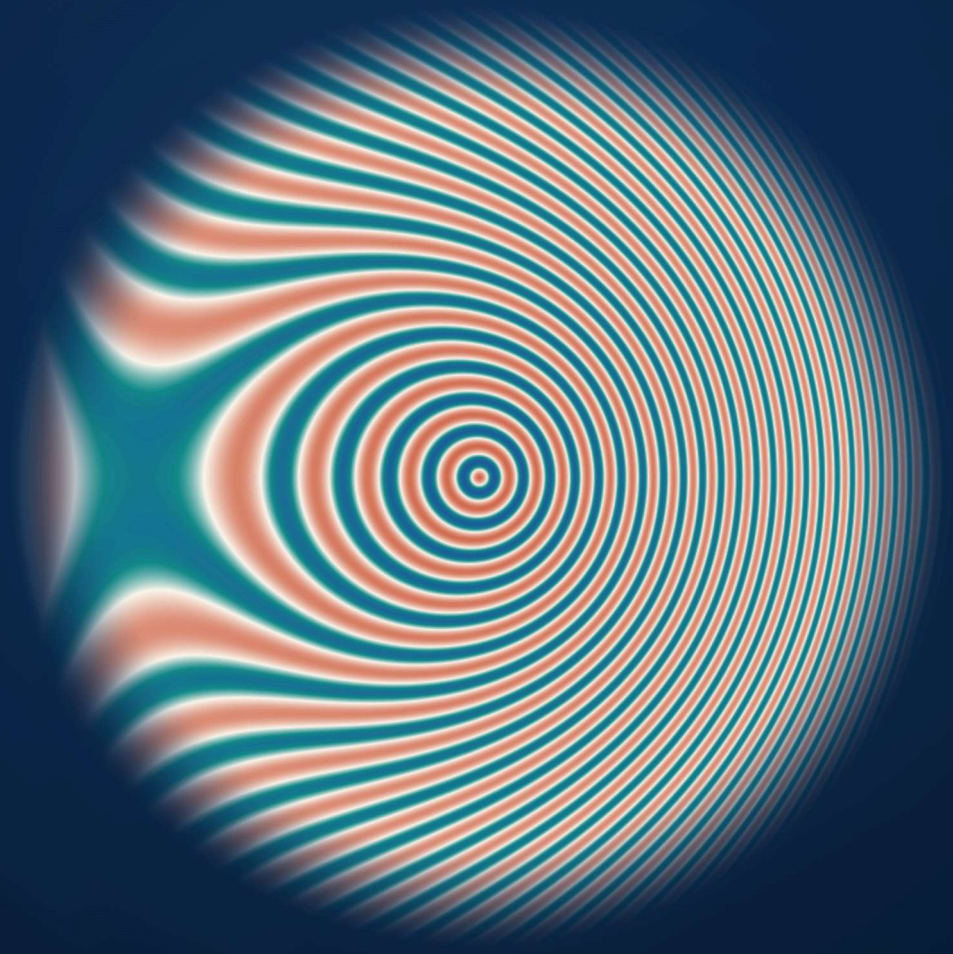
MSC THESIS · APPLIED MATHEMATICS

Frequency Transfer

Learned Frequency Operator for Optimized Helmholtz Solver Efficiency

Fenna Elise Kiewiet de Jonge



TU Delft

2026

Frequency Transfer

Learned Frequency Operator for Optimized Helmholtz Solver Efficiency

Thesis report

by

Fenna Elise Kiewiet de Jonge

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended publicly on July 15th, 2026 at 15:30 (UTC +1).

Thesis committee:	
Chair:	Prof.dr.ir. C. Vuik
Supervisors:	Prof.dr.ir. L. Demanet
	Dr.ir. V. Dwarka
External examiner:	Dr.ir. T. Nane
Place:	Faculty of EEMCS, Delft
Project Duration:	September, 2025 - July, 2026
Student number:	5160685

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Faculty of Electrical Engineering, Computer Science & Mathematics · Delft University of Technology

Preface

Dear reader, what a journey this has been.

After completing my BSc with Kees as my guide, I found myself drawn deeper into this kind of mathematics, the kind that builds smart models for problems the world actually cares about. Kees taught me not to be afraid, to go step by step, and in doing so gave me a confidence.

Curiosity has always been a thread running through everything I do. It has led me towards a rich life, shaped by a wide palette of challenges, places, hobbies, and people. Some of the clearest highlights are Delft Hyperloop 07, where I encountered the strongest *we-can-do-it* spirit I have ever seen; my first startup, ETNA Tech; sailing at the Brielse Maas, where my offshore dreams began to take shape; the Sustainability Board of the DSC; and the IDEA League.

Each of these experiences taught me something about ambition. They gave me the tools to dream big without being frightened by my own goals: to recognise intrinsic motivation, gather the right people and resources, and work with focus.

Earlier still, UWC Atlantic College planted the seed. It taught me that the world is endlessly exciting if you are willing to be a little uncomfortable in it. That discomfort, I learned, is where many interesting things begin. And through Charly's Angels, my dear peers, the world also began to feel like home.

Now one last (?) academic stretch, at the number one place in the world for mathematics: the Massachusetts Institute of Technology. Laurent showed me a very different approach: reach for the stars, be ruthlessly critical, and never lose the excitement. Follow your north star. Between the hard work in the basement, I surfaced for air at MITOC, the MIT Sailing Pavilion, and the Offshore Race Team.

To my great advisors, Kees and Laurent. To my dear club Pablo, de VB19. And to Chandler.

Dankjewel, lieve pap, mam, Mau, Huug, Flip.

And, thanks to myself. You rock.

Many doors open after this. More than excited to step into the grown-up world, and to soon sail the Atlantic. *Maak je borst maar nat!*

Abstract

This thesis studies whether information from a lower-frequency Helmholtz solve can be reused to reduce the cost of a higher-frequency solve. The approach is to learn an upward frequency-transfer operator $\mathcal{T}_{\uparrow} : u_{\omega_L} \mapsto u_{\omega_H}$ and to use the predicted high-frequency field as a warm start for CSL-preconditioned FGMRES.

The purpose of the thesis is not only to predict visually or norm-wise accurate wavefields, but to test whether such predictions are useful to the linear solver. A Krylov solver is driven by residuals, so a learned initial guess must be evaluated through the high-frequency operator A_{ω_H}, not only through its field error. The thesis therefore develops a solver-facing framework for learned frequency transfer, relating field prediction, initial-residual quality, and FGMRES convergence.

The study combines a finite-difference/PML discretisation of the Helmholtz equation, CSL-preconditioned FGMRES as the classical baseline, and a learned CNN/U-Net transfer model trained on multi-source wavefields. The method is examined through one-dimensional Dirichlet diagnostics, one-dimensional PML and two-dimensional FD/PML settings. Learned frequency transfer is treated as a warm-start mechanism and assessed by the quantities that matter to the solver: initial residuals, finite-budget residual reduction, and iteration counts.

During the work of this thesis, the author used AI-based tools (OpenAI Codex, Anthropic Claude, and Google Gemini) to assist with coding and writing. All output was reviewed and edited by the author, who takes full responsibility for the content of this work.

Contents

Nomenclature

List of Abbreviations

CNN	Convolutional Neural Network
CSL	Complex Shifted Laplacian
FD	Finite Difference
FD/PML	Finite-difference discretisation with a perfectly matched layer
FGMRES	Flexible Generalized Minimal Residual Method
FTO	Frequency Transfer Operator
FWI	Full-Waveform Inversion
GMRES	Generalized Minimal Residual Method
PDE	Partial Differential Equation
PML	Perfectly Matched Layer
PPW	Points per wavelength
RHS	Right-hand side
RMS	Root mean square
U-Net	U-shaped encoder–decoder network
UMFPACK	Unsymmetric MultiFrontal sparse direct solver

List of Symbols

β	CSL imaginary shift parameter
Δ, Δ_h	Continuous and discrete Laplacian
λ	Wavelength, $\lambda = 2\pi/k$
$\lambda_k(\omega)$	k-th 1D Dirichlet eigenvalue of the discrete Helmholtz operator
$\mathbf{x}, \mathbf{x}_s$	Spatial coordinate and source position
$\mathcal{E}(\widetilde{T})$	Intertwining defect of an approximate transfer map
$\mathcal{K}_k$	Krylov subspace of dimension k
$\mathcal{L}_\omega$	Continuous Helmholtz operator
$\mathcal{L}_{\text{field}}$	Complex relative ℓ_2 field loss
$\mathcal{L}_{\text{op}}$	Operator-aligned correction loss
$\mathcal{L}_{\text{prec}}$	Preconditioned residual loss
$\mathcal{R}_\theta, \mathcal{P}_\theta$	Learned residual restriction and correction prolongation
RelL2	Complex relative ℓ_2 field error
Ω, Ω_h	Continuous domain and finite-difference grid
$\omega, \omega_L, \omega_H$	Angular frequency; low and high frequencies
$\Omega_{\text{phys}}, \Omega_{\text{PML}}$	Physical interior and PML region
ρ	Per-sample RMS normalisation factor
σ_g	Gaussian source width
σ_j, σ_0	PML damping profile and peak damping
θ	Trainable network parameters
$\widehat{u}_{\omega_H}$	Network-predicted high-frequency wavefield
$A_\omega, A_{\omega_L}, A_{\omega_H}$	Discrete Helmholtz matrix at ω, ω_L, and ω_H
a_i, φ_i	Amplitude and phase of source component i
c, k	Wave speed and wavenumber, with $k = \omega/c$
$c_k(\cdot)$	Modal coefficient in the Dirichlet eigenbasis
$E_{\text{prec},0}$	Initial preconditioned relative residual
$E_{\text{res},0}$	Initial true relative residual
f, b	Continuous source term and discrete RHS vector
K	Number of Gaussian source components in one sample
$k^\star$	Near-zero Dirichlet mode index
M_θ^{-1}	Learned or hybrid preconditioner
M_{CSL}	CSL preconditioner
n, h	Grid points per dimension and grid spacing
n_{pml}	PML width per side
R	PML reflection coefficient
s_j	PML coordinate-stretching factor in direction j
$T_{\uparrow,\theta}$	Learned upward frequency-transfer map
$T_\uparrow^{\text{sol}}, T_\downarrow^{\text{sol}}$	Exact upward and downward solution-transfer operators
$u_\omega, u_{\omega_L}, u_{\omega_H}$	Helmholtz solution at frequency ω; low- and high-frequency solutions
v_k	k-th Dirichlet sine eigenvector
x_0, e_0, r_0	Initial guess, initial field error, and initial residual
x_k, r_k, z_k	FGMRES iterate, residual, and preconditioned direction

List of Figures

List of Tables

1 Introduction

The Helmholtz equation governs time-harmonic wave propagation, and its numerical solution at high frequency is among the long-standing hard problems of computational science. This chapter states why the problem is difficult, why it matters in practice, and why information from lower frequencies may help. The central question is not only whether a neural network can predict a high-frequency wavefield, but whether such a prediction is useful to the linear solver that ultimately controls the computation. The formal development is deferred to the chapters that follow. The purpose here is to fix the motivation, the scope, and the contributions before any technical detail is introduced.

1.1. Motivation

Solving the Helmholtz equation at high frequency is a computational core shared by many imaging and design problems. Inverse scattering formulations occur in medical imaging, sonar, radar, geophysical exploration, and nondestructive testing, where the object of interest is inferred from measured acoustic or electromagnetic wavefields [1]. Photonic and electromagnetic device design similarly relies on repeated frequency-domain field solves, often inside optimisation or inverse-design loops [2]. These applications share the same numerical difficulty at high frequency, and from this point the focus is restricted to seismic imaging, which serves as the driving application throughout this thesis.

Modern seismic reconstruction is dominated by full-waveform inversion, which fits a velocity model to observed data by repeatedly solving the forward wave problem across a band of frequencies [3]. Each forward solve in the frequency domain is a Helmholtz problem, and a single inversion requires many such solves on related media and for many right-hand sides [4]. A reduction in the cost of one high-frequency solve can therefore compound across an entire inversion, as quantified in Section 2.1.

1.2. Computational cost of high-frequency Helmholtz problems

The cost of solving the Helmholtz equation grows faster than the frequency itself. Resolving an oscillatory solution requires a fixed minimum number of grid points per wavelength, so the number of unknowns grows with frequency on a fixed domain [5]. This local condition is not sufficient, because the discrete solution also accumulates a phase error, the pollution error, that grows with frequency even when the wavelength is resolved. Suppressing this error as the frequency rises requires the grid to be refined faster than the wavelength shrinks, which drives a pollution-free discretisation to $N = \mathcal{O}(\omega^3)$ unknowns in two dimensions [6].

This $\mathcal{O}(\omega^3)$ estimate counts unknowns under refinement, not the cost of any particular solve, and it is a worst case that the present work does not pay. The experiments in this thesis hold the grid fixed and accept the growing pollution error rather than refining it away, so the problem size is constant across the frequency pairs and the ω^3 growth does not arise directly. The matrix is sparse, so the cost of a single solve is also far below the $\mathcal{O}(N^3)$ cost of a dense factorisation. With problem size held fixed, the binding cost is the number of Krylov iterations. This iteration count still grows with frequency, and reducing it is the solver-facing objective of this thesis, as developed in Chapters 3 and 4.

1.3. Failure of classical solvers at high wavenumber

Even at a fixed problem size, the resulting linear system is difficult to solve by iterative methods. Standard multigrid and domain-decomposition methods lose their usual effectiveness as the wavenumber grows, for reasons that are by now well understood [7]. The discrete Helmholtz operator is indefinite, and the absorbing boundary layer used to truncate the domain makes it non-normal [8]. Its spectrum and pseudospectrum become increasingly unfavourable as the frequency increases, so a preconditioner constructed at one frequency does not give frequency-independent convergence at higher frequencies. The Complex Shifted Laplacian (CSL) is the standard remedy [9], and combined with a multigrid hierarchy it is among the strongest classical approaches for this regime [10]. Even so, its iteration count still grows at least linearly with frequency [11]. Chapter 4 makes this baseline precise.

This observation motivates the use of additional information. In multi-frequency workflows, a high-frequency solve is not performed in isolation: lower-frequency solves in the same physical configuration have already been computed. A classical iterative method uses the high-frequency right-hand side and operator, but it does not automatically exploit the lower-frequency wavefield. Learned frequency transfer is one possible way to inject this available information into the high-frequency solve.

1.4. Why a learned method must be compared critically

A learned method is not useful merely because it predicts plausible high-frequency wavefields. Its offline cost must be justified. Data generation requires many high-fidelity solves, and the cost of those solves must either be amortised over many later right-hand sides for the same operator, or the method must target a regime where classical structure-exploiting methods are no longer available.

This distinction is important because the linear Helmholtz problem contains structure that classical numerical methods can exploit directly. In homogeneous or smoothly varying media, Green's-function representations, quadrature expansions, low-rank approximations, subspace recycling, and rational approximants can reuse information across sources or nearby frequencies. These methods use linearity and superposition explicitly; a neural model must learn such structure from data. Therefore, the present thesis does not claim that a learned warm start is a replacement for the best classical methods in the favourable linear regime.

The role of the linear FD/PML problem is more specific. It is a controlled baseline in which the relevant obstruction can be isolated. If a learned frequency-transfer map is not compatible with the residual geometry of the linear discrete Helmholtz operator, then it is unlikely to be reliable in harder settings. Conversely, understanding this obstruction is a necessary step towards hybrid learned solvers for regimes where classical linear structure is less decisive, such as non-smooth distributed sources, near-source quantities, many-parameter families, or nonlinear wave problems without a Green's function or superposition principle.

1.5. The frequency transfer idea

Solving each frequency independently discards the cross-frequency information already available in a multi-frequency workflow; this is the entry point for this thesis. Helmholtz solutions on the same medium at related frequencies share their propagation geometry, and differ by a change in amplitude scale and a phase advance [12]. In a frequency sweep, the low-frequency system is solved before the high-frequency system, so the cheaper low-frequency field is already available when the high-frequency system is assembled. The aim is to reuse this field by mapping it to an approximation of the high-frequency solution, so that the high-frequency solve begins from an informed initial guess rather than from zero.

The natural measure of such a map is its field error, and the learned map developed here is trained to minimise that error. Iterative convergence, however, is governed by the residual. If u_{ω_H} is the exact high-frequency solution and x_0 is the learned initial guess, then

$$e_0 = u_{\omega_H} - x_0, \qquad r_0 = b - A_{\omega_H} x_0 = A_{\omega_H} e_0. \tag{1.1}$$

The solver therefore does not see the field error directly. It sees the field error after it has been acted on by the high-frequency discrete operator. A small field error need not produce a small initial residual, because the operator weights different error components unevenly.

This residual elevation is best interpreted as a turn-on transient rather than as a failure of the learned field. The learned prediction may be close to the exact wavefield in an ℓ^2 sense, but the first residual evaluation tests whether the remaining error is compatible with A_{ω_H}. Components that are small in field norm can still be visible to the residual if they lie in operator-sensitive directions. The relevant conclusion is therefore not that an accurate field is a bad start, but that field accuracy alone is incomplete: a learned warm start must also be compatible with the residual seen by the discrete operator.

This distinction also separates a warm start from a preconditioner. A warm start can improve the initial state, but it does not change the operator, the preconditioner, or the asymptotic Krylov convergence mechanism. Its effect is therefore expected to be largest in the early, finite-budget part of the residual curve. A learned preconditioner would have to do more: it would have to act on residuals and corrections during the iteration. This is why the thesis treats learned upward solution transfer as the first step, and then uses the observed residual amplification to motivate residual-aware hybrid preconditioning as the natural continuation.

Chapter 5 derives the transfer operator and the error-to-residual relation. Chapter 8 then interprets the experiments through the resulting residual-amplification mechanism. This mechanism is the main analytical contribution of the thesis: it identifies when cross-frequency information remains useful after being inserted into the high-frequency Helmholtz solve.

1.6. Contributions of this thesis

This thesis makes five contributions.

- **Solver-facing formulation of learned frequency transfer.** The thesis formulates learned frequency transfer not as a stand-alone wavefield-prediction problem, but as a proposed input to a Krylov solver. It therefore distinguishes between a learned solution-transfer map, a warm start for FGMRES [13], and a genuine preconditioner that acts on residuals during the iteration.
- **Frequency-transfer model for Helmholtz warm starts.** A learned upward transfer map is constructed to predict a high-frequency wavefield from a lower-frequency solution on the same physical configuration. The predicted field is inserted as initial guess for the high-frequency CSL-preconditioned FGMRES solve, while the operator, right-hand side, and CSL preconditioner are kept fixed.
- **Residual-compatibility analysis.** The thesis identifies why field accuracy and solver usefulness are not the same criterion. In the one-dimensional Dirichlet setting, the relation $r_0 = A_{\omega_H} e_0$ can be resolved mode by mode, showing how the high-frequency operator transforms remaining field error into the residual seen by FGMRES. This provides the analytical basis for evaluating learned warm starts through both field error and residual behaviour.
- **Controlled experimental ladder.** The learned warm start is evaluated across a sequence of increasingly solver-realistic settings: one-dimensional Dirichlet diagnostics, one-dimensional PML diagnostics, and two-dimensional FD/PML experiments. This ladder is designed to separate three effects that are easily conflated: wavefield prediction quality, compatibility with the discrete operator, and actual improvement of the FGMRES residual trajectory.
- **Critical comparison with classical structure.** The thesis positions learned frequency transfer against strong classical baselines rather than against weak strawmen. In favourable linear regimes, classical methods can exploit Green's functions, superposition, low-rank structure, recycling, and CSL-based preconditioning. The learned method is therefore assessed as a possible hybrid solver component, not as a replacement for established Helmholtz methods.

The central contribution is the solver-facing interpretation of learned frequency transfer. The thesis asks whether information contained in a cheaper low-frequency solve can be inserted into a high-frequency Helmholtz solve in a way that is useful to CSL-preconditioned FGMRES. The key standard is therefore not visual quality alone, but residual compatibility with the operator A_{ω_H} and improvement of the subsequent Krylov trajectory.

The experimental results are interpreted through this standard. Field-error metrics assess whether the network has learned a meaningful cross-frequency map. Initial residuals test whether the predicted field is compatible with the high-frequency discrete equation. FGMRES curves then determine whether the

warm start lowers the final iteration count, which it does by removing error components the iteration would otherwise rebuild rather than by lowering the initial residual. This separation of field accuracy, residual compatibility, and Krylov convergence is the evaluation standard applied throughout the thesis.

Prior data-driven approaches to the Helmholtz equation learn solution maps directly [14], build learned surrogates and solvers [15], extrapolate missing frequency content for inversion [16], or embed neural components inside iterative solvers and preconditioners [17]. The present work does not claim that a new architecture alone resolves the Helmholtz problem. Instead, it isolates the condition that any learned warm start must satisfy before it can be useful to a Krylov solver: the prediction must be compatible with the residual geometry of the high-frequency operator.

The literature reviewed in this thesis was surveyed under the following primary search terms: *Helmholtz equation iterative solver*, *complex shifted Laplacian preconditioner*, *learned warm start Krylov methods*, *neural operator Helmholtz equation*, *frequency continuation full-waveform inversion*, and *neural preconditioner residual compatibility*. Coverage extends to December 2025; papers appearing after that date are not systematically included.

The remaining chapters proceed from physics to mathematics to machine learning to experiments. Chapter 2 develops the continuous physical model and the cross-frequency coherence that motivates frequency transfer. Chapter 3 constructs the finite-difference/PML operator and analyses resolution, pollution, and boundary effects. Chapter 4 reviews Krylov methods and the Complex Shifted Laplacian baseline. Chapter 5 introduces learned solution transfer as a warm-start mechanism and derives the residual-compatibility issue. Chapter 6 specifies the datasets, learned models, solver protocol, and evaluation metrics. Chapter 7 reports the one-dimensional and two-dimensional results. Chapter 8 interprets the evidence, and Chapter 9 answers the research questions and sets out future work.

Part I

Preliminary Analysis

2

Physical and Mathematical Model

High-frequency Helmholtz problems are expensive because the solution develops finer spatial scales and the operator becomes increasingly difficult to invert as frequency grows: the wavelength shrinks, the spectrum becomes indefinite, and near-zero spectral components become more prominent. After discretisation, these same facts force finer spatial resolution and lead to larger or harder linear systems. This chapter develops the continuous model that makes these statements precise. It establishes the equation, the boundary treatment, and the scale separation that determines cost, and it formulates the frequency-coherence observation that motivates the rest of the thesis. Discretisation, solver analysis, and the learned transfer operator are described in Chapters 3, 4, and 5.

This chapter develops the continuous physical model and identifies the properties that make its numerical solution expensive at high frequency. The unknown is a complex-valued field $u_\omega(\mathbf{x})$ defined on a bounded domain $\Omega \subset \mathbb{R}^2$ and acted on by differential operators. The detailed construction of the discrete operator is given in Chapter 3; this chapter refers to discrete quantities only when they are needed to state the difficulty (grid resolution, the discrete linear system) or to formulate the research question. Figure 2.1 illustrates the pipeline that links the physical system to the object studied in later chapters.

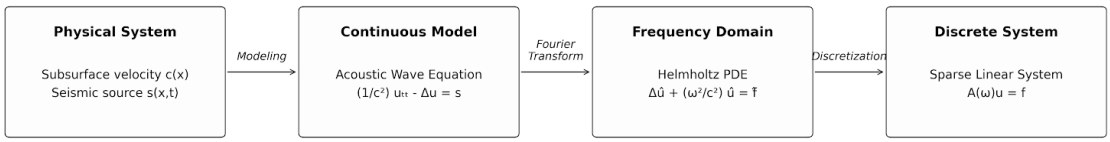


Figure 2.1: Conceptual pipeline from the physical system to the Helmholtz equation, its qualitative challenges, and the numerical treatment of Chapter 3.

2.1. Application context: seismic modelling

Wave phenomena arise in acoustics, seismology, medical imaging, and electromagnetic scattering [18]. The motivating application for this thesis is seismic modelling. Beyond the classical context of hydrocarbon exploration, accurate frequency-domain seismic modelling is important for applications such as geothermal assessment, geohazard monitoring, underground CO_2 storage, and the characterisation of subsurface reservoirs for hydrogen storage [3]. Each of these applications relies on repeated forward solves of the Helmholtz equation across a band of frequencies, so a per-solve speedup translates directly into either higher-resolution monitoring at fixed cost or lower cost at fixed resolution.

In reflection seismology, controlled acoustic sources at or near the surface emit waves that travel through the subsurface. The Earth is not homogeneous: different rock and sediment layers have different

wave speeds and densities, and geological faults can break or displace those layers. When a wave reaches such a contrast, part of its energy reflects, part refracts, and part continues deeper into the medium. Receiver arrays record the returning signals. These measurements carry information about the geometry and material properties of the subsurface.

Each shot experiment is useful only after it has been related to a simulated wavefield. Seismic imaging and inversion therefore require many forward wave-propagation solves: one simulates how a proposed subsurface model would respond to a source, compares that response with the recorded data, and then uses the mismatch to form an image or update the model. In the frequency domain, each such forward simulation becomes a Helmholtz solve, repeated over multiple frequencies. Low frequencies carry the large-scale structure of the subsurface, while higher frequencies add finer spatial detail.

Two standard examples make this repeated-solve structure explicit. Reverse-time migration forms an image by propagating source and receiver wavefields through the model [19], while full-waveform inversion (FWI) updates the model itself by repeatedly matching simulated and observed data [3]. Both workflows depend on large numbers of forward solves, and FWI additionally requires repeated adjoint-based gradient evaluations. The total cost is therefore multiplicative: a single 3D marine survey may contain 10^4 to 10^5 shots [3], frequency-domain FWI typically uses several frequency bands in a multi-scale schedule [20], and each band may require many nonlinear iterations. A complete workflow can therefore require 10^6 or more individual wave solves, so even a modest reduction in the cost of one high-frequency solve can have a large cumulative effect.

This thesis targets that repeated-solve bottleneck by reusing information between neighbouring frequency levels. The transfer pairs are $\omega_L \to \omega_H = 2\omega_L$ with $\omega_L \in \{16, 32, 64\}$. Each pair asks the same question: whether a cheaper lower-frequency solution contains enough information to provide a solver-useful warm start for the corresponding higher-frequency Helmholtz solve.

2.2. Modelling scope and assumptions

The continuous model retains only the mechanisms that govern high-frequency numerical behaviour: oscillation, reflection, interference, and the need for absorbing boundaries. Table 2.1 maps each assumption to a single, traceable consequence; the prose that follows expands on the consequences that are not self-evident from the table alone.

Three of the assumptions in Table 2.1 carry specific consequences for the experimental design and require further explanation here, even though their physical content is standard.

The first is linear acoustics. Linear acoustics describes the regime in which pressure amplitudes are sufficiently small that the governing equation is linear in the field variable p; nonlinear effects such as harmonic generation and shock formation are therefore absent. The linearity assumption is uncontroversial in the small-amplitude regime considered here, but it is recorded explicitly because it is the basis for the data-generation strategy adopted in Chapter 6. Because $\mathcal{L}_\omega$ is a linear operator, the solution map $f_\omega \mapsto \mathcal{L}_\omega^{-1} f_\omega$ is linear in the source term. Equivalently, if $\mathcal{L}_\omega u_i = f_i$ for each index i, then the principle of superposition gives

$$\mathcal{L}_\omega \left(\sum_i u_i \right) = \sum_i f_i \,, \tag{2.1}$$

so solving once with the composite source $f = \sum_i f_i$ yields the field $u = \sum_i u_i$, the exact sum of the individual single-source responses. Each training sample is therefore constructed by summing 3–6 randomly placed Gaussian sources into a single right-hand side and solving at both ω_L and ω_H, producing physically consistent input–output pairs that expose the network to multi-source interference without requiring a separate solve per source.

The second is the source representation. The Gaussian profiles used throughout this thesis are not compactly supported in the strict mathematical sense; they decay as $e^{-r^2/(2\sigma_g^2)}$ with $\sigma_g = 2$ grid cells. Truncation at $4\sigma_g$ is applied as a practical approximation. At the truncation boundary $r = 4\sigma_g$, the pointwise amplitude falls to $e^{-8} \approx 3.4 \times 10^{-4}$ of its peak value. The fraction of one-dimensional Gaussian mass outside $\pm 4\sigma_g$, measured by the complementary error function, is $\operatorname{erfc}(4/\sqrt{2}) \approx 6.3 \times 10^{-5}$;the discarded tail therefore lies at the level of finite-difference truncation error and well below the dynamic range of the wavefield.

Table 2.1: Modelling assumptions, their physical purpose, and implications for numerical treatment. Each row maps a single physical assumption to a single consequence; cost estimates are referenced where they first appear in later chapters.

Assumption	Physical purpose	Implication for modelling
Time-stationary medium	Material properties do not vary in time	Time dependence decouples by Fourier transform; each ω admits an independent frequency-domain problem.
Linear acoustics	Wave amplitudes remain in the linear regime	Solutions superpose over sources; multi-source training samples are formed by summing 3–6 single-source fields and solving once per pair (ω_L, ω_H).
Isotropic medium	No directional dependence in wave speed	Wavenumber reduces to the scalar $k = \omega/c$; the operator is a scalar PDE rather than a tensor system. In elastic wave propagation, by contrast, the displacement field couples to stress and strain through a rank-4 stiffness tensor, yielding a system of coupled PDEs.
Constant wave speed $c \equiv 1$	Simplifies data generation and transfer structure	Free-space Green's function is explicit (equation (2.2)); $k = \omega$ is spatially uniform across Ω.
Bounded $\Omega \subset \mathbb{R}^2$	Makes computation finite	Necessitates artificial boundary conditions (Section 2.4); the cost consequences of grid resolution and direct factorisation enter only after discretisation and are quantified in Chapter 3.
Acoustic (scalar) waves	Reduces model complexity	Avoids elastic coupling and mode conversion.
Smooth, rapidly decaying sources in Ω_{int}	Sources placed away from the absorbing layer	Gaussian profiles ($\sigma_g = 2$ cells, truncated at $4\sigma_g$); separation from the PML keeps the source effectively zero in the boundary region.
Absorbing boundaries (PML)	Emulates open-domain radiation	Introduces complex-valued coordinate stretching; the discrete operator becomes non-Hermitian.

The third assumption is the constant wave speed $c \equiv 1$. This restriction governs three downstream choices. First, the wavenumber satisfies $k = \omega$ and is spatially uniform, so u_ω oscillates at a single global spatial scale rather than at a locally varying scale set by the medium. Second, the free-space Green's function of the two-dimensional Helmholtz equation takes the explicit form

$$G_\omega(\mathbf{x}) \;=\; \frac{i}{4}\, H_0^{(1)}\big(\omega\,|\mathbf{x}|\big), \tag{2.2}$$

where $H_0^{(1)}$ denotes the Hankel function of the first kind and order zero [21].

Because the solution satisfies $u_\omega = G_\omega * f_\omega$, data generation can exploit FFT-based convolution with the Hankel kernel equation (2.2) instead of requiring a sparse direct solve for every training sample; the discrete convolution procedure is described in Chapter 6. Third, the frequency-transfer structure is more predictable: the amplitude envelope of u_ω is governed by source geometry and geometric spreading, rather than by medium-induced refraction, scattering, or focusing.

The constant-speed setting is therefore the natural first test of the frequency-transfer hypothesis. It isolates the question of whether a learned operator can transfer phase and amplitude information across frequencies before the additional difficulty of a spatially varying medium is introduced. The setting is not entirely free of spatial variation: the PML of Section 2.4 already equips the operator with spatially varying complex coefficients in the boundary collar, so the learned map must already accommodate a non-constant-coefficient operator near the domain edge. What is held constant is the slowness m of the physical interior, and with it the medium-induced refraction, scattering, and focusing that a heterogeneous model would introduce.

The constant-slowness assumption is therefore a principal limitation of the present work. Realistic seismic and subsurface storage applications involve strongly heterogeneous media, where the free-space convolution structure no longer applies and the transfer map must account for medium-dependent ray bending, scattering, and amplitude changes. Whether the learned map extends to such media is not established here, because the heterogeneous case has not been investigated.

Because the discrete operator already carries spatially varying complex coefficients in the PML collar, there is reason to expect the method to extend to mildly heterogeneous interiors; that expectation is positive but cannot be claimed without direct investigation.

The homogeneous transfer picture developed in Chapter 5 should therefore be read as a controlled baseline whose reach beyond constant slowness is left open. Extending the method to heterogeneous media is one of the principal future-work directions identified in Chapter 9.

Together, these assumptions yield a well-posed model that nevertheless retains the principal numerical difficulties of the Helmholtz equation: indefiniteness, non-Hermitian structure, sensitivity to grid resolution, and the need for accurate absorbing boundary conditions.

2.3. From the acoustic wave equation to the Helmholtz equation

With the modelling scope fixed, the equation governing u_ω can now be derived. The acoustic wave equation describes the evolution of pressure $p(t, \mathbf{x})$ in time t and space $\mathbf{x} \in \Omega$,

$$\frac{\partial^2 p}{\partial t^2} - c^2(\mathbf{x})\,\Delta p = s(t, \mathbf{x}), \tag{2.3}$$

where $c(\mathbf{x})$ is the (in general) spatially varying wave speed and s is a source term. The Laplacian $\Delta p = \sum_{j=1}^{2} \partial^2 p / \partial x_j^2$ measures local curvature and models how waves spread as they propagate.

For a source oscillating at a single angular frequency $\omega > 0$, both the source and the pressure field are written as

$$p(t, \mathbf{x}) = \mathrm{Re}\big\{u_\omega(\mathbf{x})\, e^{-i\omega t}\big\}, \qquad s(t, \mathbf{x}) = \mathrm{Re}\big\{f_\omega(\mathbf{x})\, e^{-i\omega t}\big\}. \tag{2.4}$$

Substituting (2.4) into (2.3) and cancelling the common factor $e^{-i\omega t}$ yields the Helmholtz equation

$$\mathcal{L}_\omega u_\omega = f_\omega, \qquad \mathcal{L}_\omega := -\Delta - \omega^2 m(\mathbf{x}), \qquad m(\mathbf{x}) = \frac{1}{c^2(\mathbf{x})}, \tag{2.5}$$

where $m(\mathbf{x})$ is the slowness of the medium. The time dependence is eliminated; the unknown is the complex-valued spatial amplitude u_ω.

The operator $\mathcal{L}_\omega$ contains two competing terms. The Laplacian $-\Delta u_\omega$ is a smoothing term: it resists sharp spatial variations and drives the field toward local averages. The mass term $-\omega^2 m(\mathbf{x})\, u_\omega$ enforces local oscillation at a rate set by the medium. In a homogeneous medium with constant c, the elementary solutions of the homogeneous equation are plane waves $u_\omega(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}}$, and substitution gives the dispersion relation $|\mathbf{k}|^2 = \omega^2 c^{-2}$: the wavenumber magnitude is fixed by ω and c, while the propagation direction is free. The associated wavelength

$$\lambda = \frac{2\pi}{|\mathbf{k}|} = \frac{2\pi c}{\omega} \tag{2.6}$$

determines the spatial scale of oscillation: higher frequencies correspond to shorter wavelengths and finer spatial structure.

The interaction between the diffusive Laplacian and the oscillatory mass term gives rise to the wave phenomena observed in practice. Diffraction occurs when waves spread after passing obstacles or openings. Interference arises when contributions from different paths overlap. Focusing occurs when variations in $c(\mathbf{x})$ act as lenses, concentrating wave energy. All of these behaviours are encoded in equation (2.5).

2.4. Open space on a bounded domain

The Helmholtz equation derived above is posed on free space, but numerical simulation must take place on a finite domain Ω. Restricting the unbounded free-space problem to a finite computational domain is referred to as domain truncation, and it is the step that introduces the artificial boundary $\partial\Omega$. Without special treatment, waves reach $\partial\Omega$ and reflect back into the interior, contaminating the solution. The boundary conditions must therefore allow outgoing waves to leave Ω without producing artificial reflections. Figure 2.2 shows the two strategies used in this thesis.

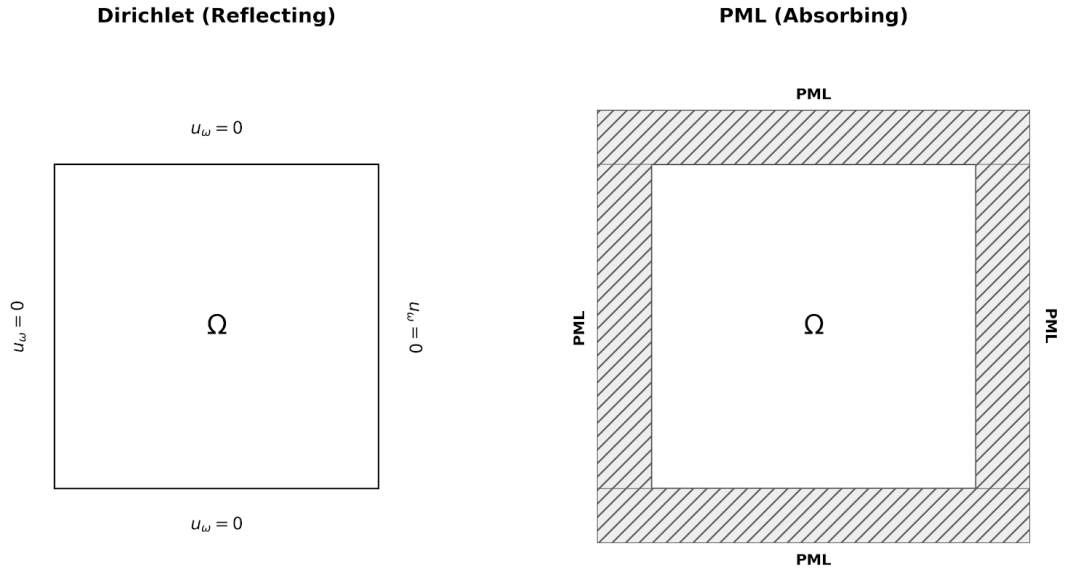


Figure 2.2: Two boundary strategies for a bounded computational domain. Left: Dirichlet wall, fully reflecting. Right: perfectly matched layer, designed to absorb outgoing waves.

The simplest condition sets the field to zero on $\partial\Omega$,

$$u_\omega = 0 \quad \text{on } \partial\Omega. \tag{2.7}$$

This Dirichlet condition is easy to implement and used as a reference case. It behaves as a perfectly reflecting wall: an outgoing wave reflects back with a phase reversal, producing artificial standing waves in the interior.

The perfectly matched layer (PML) is the absorbing strategy used throughout this thesis. A PML surrounds the physical domain with an artificial collar in which outgoing waves are exponentially damped before they can reflect [22]. The mechanism is *complex coordinate stretching*: inside the collar, the real coordinate x_j along axis j is replaced by a complex-valued counterpart $\tilde{x}_j$ defined through the stretching function

$$s_j(x_j) = \frac{\mathrm{d}\tilde{x}_j}{\mathrm{d}x_j} = 1 + \frac{i\,\sigma_j(x_j)}{\omega}, \tag{2.8}$$

where $\sigma_j(x_j) \geq 0$ is the damping profile. The ratio s_j is the complex coordinate increment per unit real increment: a positive imaginary part modifies the effective wavenumber so that waves accumulate exponential decay as they travel outward. The profile σ_j grows smoothly from zero at the interior–PML interface to a maximum value at the outer boundary, typically as a polynomial ramp.

The continuity of σ_j at the interface is what makes the layer reflectionless. Acoustic impedance characterises the resistance that a medium presents to wave propagation, and a jump in impedance across an interface causes part of an incident wave to reflect, exactly as light reflects at a glass surface. The PML introduces damping through coordinate stretching rather than through a change in the material properties of the medium, so the impedance of the collar matches that of the interior at the interface. Because σ_j starts at zero and grows continuously, no impedance jump occurs at the interface, and an outgoing wave enters the collar without reflection. In the continuous setting this holds in principle at any angle of incidence and any frequency; discretisation introduces small residual reflections, whose size is controlled by the design of σ_j and quantified through the reflection coefficient below.

The key design parameter is the reflection coefficient R: the fraction of wave amplitude that returns to the interior after traversing the PML. For a collar of thickness δ with a polynomial damping profile,

$$R = \exp\left(-2\int_0^{\delta} \sigma(x)\,\mathrm{d}x\right). \tag{2.9}$$

The PML is $n_{\mathrm{pml}} = 112$ grid cells thick on each side, and σ_0 is calibrated as a function of ω to maintain $R \leq 10^{-2}$ across all frequencies. The precise scaling and its derivation are given in Chapter 3.

2.5. Qualitative properties of the Helmholtz operator

With the equation and its boundary treatment in place, the next question is what makes the resulting operator hard to invert. The structure of $\mathcal{L}_\omega = -\Delta - \omega^2 I$ determines the numerical and algebraic difficulty of the resulting linear system. In this context, difficulty refers specifically to the spectral properties of the operator that impede the convergence of iterative solvers. Three properties are primarily responsible for this behaviour: indefiniteness, non-Hermitian structure, and near-zero spectral components.

The first is **indefiniteness**. An operator is coercive if it satisfies $\mathrm{Re}\langle \mathcal{L}u, u\rangle \geq \alpha\|u\|^2$ for some $\alpha > 0$ uniformly in u. While the Laplacian $-\Delta$ is coercive under standard boundary conditions, the mass term $-\omega^2 I$ acts in opposition. For small ω the Laplacian dominates and $\mathcal{L}_\omega$ remains coercive. As ω grows, the mass term shifts the spectrum across the origin. Consequently, the operator becomes indefinite, its spectrum contains eigenvalues on both sides of the imaginary axis, and coercivity is lost. Solvers relying on positive definiteness, such as the Conjugate Gradient method, fail entirely. Krylov methods like GMRES can handle indefinite systems but face significant convergence bottlenecks as the spectrum encloses the origin [9].

The second is **non-Hermitian behaviour**. In a lossless medium, $\mathcal{L}_\omega$ is self-adjoint. The perfectly matched layer introduces complex-valued coordinate stretching to emulate open-domain radiation, which renders the operator non-self-adjoint. The practical consequence concerns how solver convergence should be predicted. For a normal operator, the convergence of a Krylov method such as GMRES is governed by the distribution of the eigenvalues alone. Loss of normality breaks this link, because the eigenvectors need not be orthogonal and can form an ill-conditioned basis. In that regime a favourable eigenvalue distribution can coexist with slow convergence, so the eigenvalues alone become an unreliable indicator of solver difficulty. Sharper bounds are then obtained from the field of values or the pseudospectrum, which account for non-normality directly [8].

The third property is the emergence of **near-zero spectral components**. As ω increases, the shifted operator $-\Delta - \omega^2 I$ moves more Laplacian eigenvalues close to zero. These components are not problematic because of "resonance" in a vague physical sense, but because they make the discrete operator nearly singular in selected directions. Equivalently, there exist modes for which the positive Laplacian contribution and the negative mass shift almost balance. These near-zero eigenvalues increase the sensitivity of the linear system and create the residual-amplification mechanism analysed in Chapter 5. In the PML case the same idea survives in a non-normal form: the relevant difficulty is no longer captured by real eigenvalues alone, but by near-null directions and the pseudospectral sensitivity of the full finite-difference and PML operator.

2.6. Frequency, wavelength, and the multiscale challenge

The three operator-level properties above all share a single physical origin: as ω grows, the wavelength $\lambda = 2\pi/\omega$ shrinks, so the solution u_ω varies on increasingly fine spatial scales. Resolving these oscillations on a finite grid is the source of every discrete cost estimate in this thesis, but the resolution question itself, namely how fine a grid is required and how the global solution error grows with ω, depends on the choice of discretisation. Chapter 3 carries out that analysis in full and quantifies both the local sampling condition (the points-per-wavelength heuristic) and the global pollution error that remains even when the local condition is satisfied. The remainder of the present chapter therefore returns to the continuous model and formalises the cross-frequency coherence that motivates the rest of the thesis.

2.7. Research question and scope

The properties developed in the preceding sections, indefiniteness, non-Hermitian structure, condition-number growth, and resolution sensitivity, together explain why high-frequency Helmholtz problems are expensive. Classical preconditioners such as the Complex Shifted Laplacian (CSL) [23] improve Krylov convergence, but their effectiveness still deteriorates as the frequency ω increases. A complementary observation motivates the present approach: CSL is effective on the bulk of the spectrum but loses effectiveness on the near-zero components identified in Section 2.5, which is precisely where a learned transfer map can contribute.A further limitation is that each frequency is usually treated as an independent linear system, even though wavefields generated in the same medium at related frequencies contain strongly correlated information.

The broader goal of this thesis is to exploit this cross-frequency structure for faster high-frequency Helmholtz solves. For a fixed medium, the operator $\mathcal{L}_\omega$ varies systematically with ω. Provided ω remains bounded away from the (complex) eigenvalues of $\mathcal{L}_\omega$, which the PML displaces off the real axis by construction, the corresponding fields evolve coherently with frequency [12]. The coherence is sharper than a qualitative statement: for a factor-of-two frequency pair $(\omega_L, \omega_H) = (\omega/2, \omega)$ in two dimensions, the asymptotic Hankel expansion predicts that the high-frequency wavefield retains the spatial amplitude pattern of the low-frequency wavefield, scaled globally by $\sqrt{\omega_L/\omega_H} = 1/\sqrt{2} \approx 0.707$, while the phase advances by a deterministic amount along each propagation path. A solution computed at a cheaper lower frequency therefore contains structured, quantifiable information about the corresponding higher-frequency solution; Chapter 5 derives this prediction in detail and reports the empirical ratio measured on the generated dataset.

This idea is expressed through a frequency-transfer operator $\mathcal{T}_{\omega_L \to \omega_H}$, with $\omega_L < \omega_H$, such that

$$u_{\omega_H} \;\approx\; \mathcal{T}_{\omega_L \to \omega_H} u_{\omega_L}. \tag{2.10}$$

At the operator level, the structural target for an ideal pair of transfer maps is that the high-frequency inverse should factor through the low-frequency inverse, with $\mathcal{T}_{\omega_H \to \omega_L}$ and $\mathcal{T}_{\omega_L \to \omega_H}$ playing the respective roles of restriction and prolongation. The corresponding heuristic relation is

$$\mathcal{L}_{\omega_H}^{-1} \;\approx\; \mathcal{T}_{\omega_L \to \omega_H}\, \mathcal{L}_{\omega_L}^{-1}\, \mathcal{T}_{\omega_H \to \omega_L}. \tag{2.11}$$

This is not a formal identity. Equation (2.11) is exact only in idealised settings: free space, a homogeneous medium, and an untruncated domain that carries no artificial boundary (the truncation introduced in Section 2.4). It degrades as soon as a PML is introduced or the medium becomes heterogeneous. It is therefore useful as motivation rather than as a property the thesis claims.

The single pair (ω_L, ω_H) used throughout this thesis is the two-level instance of a more general frequency hierarchy $\omega_0 < \omega_1 < \cdots < \omega_J$. Replacing the pair by adjacent levels $(\omega_\ell, \omega_{\ell+1})$ turns the restriction and prolongation maps $\mathcal{T}_{\omega_{\ell+1}\to\omega_\ell}$ and $\mathcal{T}_{\omega_\ell\to\omega_{\ell+1}}$ into the inter-level transfers of a frequency multigrid V-cycle, with equation (2.11) playing the role of the two-grid identity. The present work studies only the two-level case; the multilevel cascade is one of the future-work directions identified in Chapter 9.

The deviation from this ideal is, however, an exact and computable object. For any approximate map $\widetilde{\mathcal{T}}$, define the *intertwining defect*

$$\mathcal{E}(\widetilde{\mathcal{T}}) \;:=\; \mathcal{L}_{\omega_L}\,\widetilde{\mathcal{T}} \;-\; \widetilde{\mathcal{T}}\,\mathcal{L}_{\omega_H}. \tag{2.12}$$

The defect vanishes when $\widetilde{\mathcal{T}}$ satisfies the intertwining relation exactly and grows with the approximation error otherwise. Its operator norm bounds the deviation from the ideal case in which $\widetilde{\mathcal{T}}$ acts as a preconditioner, making it the natural quality measure for learned transfer operators. Chapter 5 makes this relation precise on the finite-difference grid and uses it to motivate the warm-start formulation and the preconditioning extensions discussed later.

The present thesis takes a deliberately narrower first step. Rather than constructing a full learned preconditioner, it studies whether a learned upward frequency-transfer map can provide a useful *warm start* for a CSL-preconditioned FGMRES solve. Given a low-frequency solution u_{ω_L}, the model predicts

$$\hat{x}_0 = \mathcal{T}_{\omega_L\to\omega_H} u_{\omega_L}, \tag{2.13}$$

which is used as the initial iterate for the high-frequency linear system

$$A_{\omega_H} x = b_{\omega_H}. \tag{2.14}$$

The relevant measure of success is therefore not only whether $\hat{x}_0$ is close to u_{ω_H} in a field norm, but whether it accelerates the iterative solver. The binding measure of acceleration is the number of FGMRES iterations required to reach a prescribed tolerance; wall-clock time is deliberately not used as the outcome measure, for the cost-accounting reasons set out in Section 6.6.2. The initial relative residual

$$\frac{\|b_{\omega_H} - A_{\omega_H}\hat{x}_0\|}{\|b_{\omega_H}\|} \tag{2.15}$$

is a natural intermediate quantity: it is the first thing the solver sees and is more closely aligned with the operator than the field error. It is, however, an iteration-zero quantity, and its importance is easily overstated. The elevated residual reflects how the high-frequency operator weights the remaining field error at the very first evaluation, before any preconditioning has acted. Once the CSL preconditioner and the first few FGMRES steps are applied, the operator-sensitive components that inflate this residual are damped quickly, so an initially large residual becomes far less consequential a few iterations in. What persists, and therefore controls the total iteration count, is whether the warm start has removed error components that the iteration would otherwise have had to rebuild; those components do not reappear. A warm start can thus begin from an elevated initial residual and still converge in fewer iterations. Whether the initial residual is nonetheless a more reliable predictor of iteration savings than the field error is precisely the question raised by SQ2 below.

The central research question of this thesis is therefore

> *Can learned frequency transfer produce solver-useful warm starts for CSL-preconditioned FGMRES applied to high-frequency Helmholtz problems?*

This question is refined into the following sub-questions:

> *SQ1. How accurately can an upward transfer model predict the high-frequency solution from a lower-frequency solution, and how does this accuracy degrade across the frequency pairs* $(16, 32)$, $(32, 64)$, *and* $(64, 128)$*?*
>
> *SQ2. Is field accuracy alone a reliable predictor of solver acceleration, or are PDE-aligned quantities such as the initial residual more strongly correlated with FGMRES iteration savings?*

The experimental scope is organised around these two sub-questions.

The first group of experiments addresses SQ1. A baseline learned warm-start operator $u_{\omega_L} \mapsto u_{\omega_H}$ is established and evaluated across the three pairs $(16, 32)$, $(32, 64)$, and $(64, 128)$. Whether prediction quality is limited by training time, model capacity, or architecture is studied by varying each factor in turn. These runs are reported using both prediction metrics, such as the relative field error, and solver metrics, such as the initial residual, the FGMRES iteration count, the convergence history, and the solve time. Reporting both classes of metric on the same runs is what makes the comparison required by SQ2 possible.

The second group of experiments addresses SQ2 directly, by asking whether field accuracy is a reliable predictor of solver acceleration. A field-accurate warm start can still produce a large initial residual, because the discrete operator amplifies precisely the near-zero-eigenvalue components that the network fails to suppress. This is the residual-amplification mechanism identified in Section 2.5 and analysed quantitatively in Chapter 5.

Two diagnostic probes would sharpen the question further. The first is a systematic correlation of relative field error and initial residual against realised FGMRES iteration savings across all runs. The second augments the supervised loss with a residual penalty involving $A_\omega \hat{x}_0 - b_\omega$, so that the prediction is encouraged to satisfy the discrete high-frequency equation rather than only to match the field. The present experiments answer SQ2 through same-run comparisons of field error, true residual, and preconditioned residual on identical test problems (Chapter 7); the correlation sweep and the residual-penalty objective are identified as the highest-priority follow-up and are specified in Chapter 9.

Second, the supervised loss is augmented with a residual penalty involving $A_\omega \hat{x}_0 - b_\omega$, so that the prediction is encouraged to satisfy the discrete high-frequency equation rather than only to match the field. The field-error objective is retained as the primary training target throughout; the residual penalty is examined only as a diagnostic that tests whether aligning the loss with the Krylov objective closes the gap between field accuracy and solver acceleration.

A supporting control isolates true frequency transfer from source-to-solution synthesis. Because the target solution satisfies $A_\omega u_\omega = b_\omega$, the source term may carry information that is not recoverable from u_{ω_L} alone. Comparing a model that receives only u_{ω_L} with a model that additionally receives the source separates the contribution of genuine cross-frequency transfer from that of source-to-solution mapping; the solver-facing two-dimensional model is deliberately denied the source for this reason (Section 6.5.4).

Two directions are retained as natural extensions, to be pursued once the warm-start mechanism is understood together with the conditions under which it does not accelerate the solve. Two such conditions are identified in this thesis. The first is that a field-accurate prediction can still leave a large initial residual, because the high-frequency operator amplifies precisely the near-zero-eigenvalue components that the network does not suppress, so accuracy in the field norm does not guarantee a small residual. The second is that a reduced initial residual need not translate into fewer iterations, because the components the warm start removes are not necessarily the ones that govern the slow part of the Krylov convergence; an informed start that lowers the residual on already-fast directions buys little. Understanding both conditions is what the two extensions below are designed to address.

The first is a learned preconditioner that acts on the Krylov residual at every iteration rather than supplying a single initial guess. The intertwining defect $\mathcal{E}(\widetilde{\mathcal{T}})$ of equation (2.12) motivates this directly: driving the defect toward zero turns $\widetilde{\mathcal{T}}$ into an operator that maps a residual r of the high-frequency system to a consistent correction $z \approx A_{\omega_H}^{-1} r$, which is the defining action of a preconditioner. The second is a recursive multilevel cascade across the frequency hierarchy $\omega_0 < \cdots < \omega_J$: each adjacent transfer $\omega_\ell \to \omega_{\ell+1}$ supplies the warm start for the next level, so that a high-frequency target is reached through a sequence of cheap single-octave transfers rather than a single large jump, with equation (2.11) acting as the two-grid identity at each level. Both directions are developed in Chapter 9.

Although the Helmholtz equation and its discrete operator are linear in the field for a fixed frequency and medium, the solver framework is chosen to admit preconditioned actions that may vary between iterations. For this reason the solver used here is FGMRES rather than standard GMRES. FGMRES contains fixed-preconditioner GMRES as a special case, while also admitting the learned or adaptive preconditioning actions required by the extensions above.

All experiments in the present scope are conducted in two spatial dimensions on a unit square domain with constant wave speed $c \equiv 1$. The main transfer pairs are

$$(\omega_L, \omega_H) \in \{(16, 32), (32, 64), (64, 128)\},$$

so that each transfer spans a factor of two. The factor-of-two choice is deliberate. It mirrors the frequency cascades used in multi-scale full-waveform inversion [4], where each band roughly doubles the previous one. It is also the smallest frequency gap at which the high-frequency solve is meaningfully more expensive than the low-frequency solve: smaller ratios produce nearly identical solver costs and offer no target to accelerate. This single-level, factor-of-two setting is therefore the simplest non-trivial test of frequency transfer, since it asks whether a cheaper half-frequency solve provides enough information to accelerate the corresponding full-frequency solve.

This chapter has established the physical and mathematical setting: the acoustic wave equation, its time-harmonic reduction to the Helmholtz equation, boundary modelling through the PML, the numerical difficulties of the discrete operator, and the frequency-coherence observation that motivates learned frequency transfer. Chapter 3 constructs the finite-difference and PML operator A_ω used by the solver. Chapter 4 then defines the CSL-preconditioned FGMRES baseline, and Chapter 5 introduces learned frequency transfer as a warm-start mechanism, with residual-aware preconditioning retained as the natural extension.

3

Numerical Discretisation

This chapter develops the right-hand box of Figure 2.1 (the discrete system $A_\omega u = b$) and verifies that each of the qualitative obstacles identified in Sections 2.5 and 2.6 reappears at the algebraic level. Three parameters are fixed: a second-order five-point stencil for the Laplacian, a fixed-width Perfectly Matched Layer (PML) with frequency-calibrated damping, and a uniform 512×512 grid shared across all frequencies. These choices isolate the effect of frequency ω on solver complexity by holding the number of degrees of freedom and the spatial geometry constant throughout the thesis.

Three design choices in this chapter together control the numerical error of the discrete Helmholtz operator. The grid resolution λ/h is held above the standard ten-points-per-wavelength threshold so that the local truncation error of the five-point stencil is bounded uniformly across the four operating frequencies (Section 3.5). The PML width $n_{\text{pml}} = 112$ and the frequency-calibrated peak damping $\sigma_0(\omega)$ together hold the boundary reflection coefficient at $R \leq 10^{-2}$ across the same frequencies (Section 3.7). The choice to fix n rather than h is then explained as a deliberate trade-off: it leaves the global pollution error free to grow with ω, but in exchange it keeps the spatial layout, the network input dimension, and the PML geometry constant across frequency pairs. The pollution growth is quantified in Section 3.6.

3.1. Continuous problem formulation

This chapter is the point at which the continuous model of Chapter 2 becomes a finite-dimensional linear system, so the notational convention used from here onward is stated once. Calligraphic operators such as $\mathcal{L}_\omega$, and functions carrying a spatial argument such as $u_\omega(\mathbf{x})$ and $f(\mathbf{x})$, denote continuous objects on Ω. Roman matrices such as A_ω and M_{CSL}, and indexed or plain vectors such as u, b, and r_0, denote their finite-difference discretisations on the grid Ω_h. The subscript h, as in Δ_h, marks a discrete differential operator. Operators and vectors are set in upright (non-bold) type throughout; boldface is not used to mark discreteness, since the calligraphic-versus-roman distinction already carries that information. Spectral quantities (eigenvalues, singular values, pseudospectra, and condition numbers) always refer to the discrete matrix A_ω unless the continuous operator $\mathcal{L}_\omega$ is named explicitly.

The starting point is the time-harmonic scalar Helmholtz equation on a bounded domain $\Omega \subset \mathbb{R}^2$:

$$-\Delta u(\mathbf{x}) - \frac{\omega^2}{c^2}\, u(\mathbf{x}) = f(\mathbf{x}), \qquad \mathbf{x} \in \Omega. \tag{3.1}$$

The wave speed is fixed at $c = 1$ throughout this thesis, as specified in Section 2.2. Outgoing radiation is enforced by the perfectly matched layer introduced in Section 2.4 and discretised in Section 3.7. In the numerical experiments, the computational domain is the unit square $\Omega = [0, 1]^2$, discretised by $n = 512$ cells per side so that $h = 1/512$. An absorbing PML collar of $n_{\text{pml}} = 112$ grid cells on each side occupies the outer frame of this grid, leaving the physical interior Ω_{phys} as the central 288×288 block, $\Omega_{\text{phys}} = [\, n_{\text{pml}}h,\ 1 - n_{\text{pml}}h\,]^2 = [0.219,\ 0.781]^2$. The continuous right-hand side f is sampled on the finite-difference grid to form the discrete vector b. The discretisation of Equation (3.1) is carried out in Section 3.2.

3.2. Finite difference approximation

The domain $\Omega \subset \mathbb{R}^2$ is discretised using a uniform Cartesian grid Ω_h with mesh spacing h. A grid point is indexed by (i, j) and represents the value $u_{i,j} \approx u(ih, jh)$. The Laplacian is approximated by the standard

second-order central difference stencil:

$$\left[\nabla^2 u\right]_{i,j} \approx \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h^2}, \tag{3.2}$$

which couples five grid points and maintains a local truncation error of $\mathcal{O}(h^2)$. This follows from the Taylor expansion of $u(x \pm h)$, where the summation of the forward and backward expansions cancels the odd-order derivatives, leaving the second-order derivative with an $\mathcal{O}(h^2)$ remainder.

While this stencil applies to all interior grid points, the weights within the absorbing boundary layer are modified by complex coordinate stretching as detailed in Section 3.7. The second-order five-point stencil is chosen over higher-order or compact alternatives for three reasons. It preserves the standard five-point sparsity pattern of A_ω, which keeps the matrix-vector product cost strictly $\mathcal{O}(N)$ and matches the local receptive-field width assumed by the convolutional architectures of Chapter 5, where each output channel depends on a spatially local patch of the input whose width must be at least as wide as the stencil support. It is the discretisation underlying the shifted-Laplacian literature that defines the CSL preconditioner used here [9], so the discrete operator is the standard one rather than a variant that would confound the comparison. Most importantly for the analysis that follows, the second-order Dirichlet operator admits the closed-form sine eigenbasis used in Proposition 5.1, whose eigenvalues already appear in Equation (3.6); a compact or higher-order stencil would perturb those eigenpairs and remove the exact modal diagnostic on which the residual-amplification analysis depends. Higher-order stencils suppress the leading pollution term but widen the stencil support and complicate the Krylov benchmarks; they are recorded as future work.

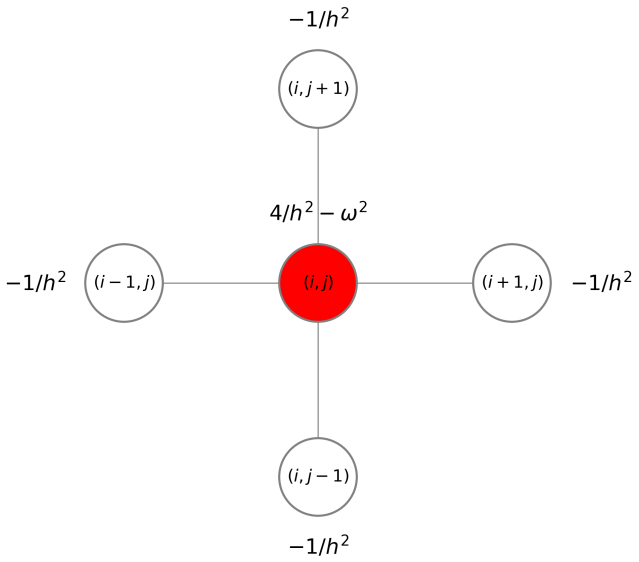


Figure 3.1: Five-point stencil for the discrete Helmholtz operator $A_\omega = -\Delta_h - \omega^2 I$. The four neighbouring grid points carry coefficient $-1/h^2$, while the central coefficient is $4/h^2 - \omega^2$. The mass shift ω^2 is the term that distinguishes the Helmholtz stencil from the Laplace stencil.

3.3. Discrete Helmholtz operator

The grid points are ordered lexicographically (row by row, left to right, top to bottom) into a single vector $u \in \mathbb{C}^N$, where $N = n^2$ is the total number of degrees of freedom and $n = 512$. Applying the approximation

in (3.2) yields the linear system:

$$A_\omega\, u = b, \tag{3.3}$$

where $b \in \mathbb{C}^N$ is the discretised source: the sampled values of the continuous source term f on the grid, written as a column vector under the same lexicographic ordering as u. The continuous source f and its discretisation b are distinguished throughout the thesis: f is a function on Ω, b is a vector in $\mathbb{C}^N$, and the relation between them is sampling on the finite-difference grid.

Each row of A_ω contains at most five non-zero entries, resulting in a sparse system with $\mathcal{O}(N)$ non-zeros. For $n = 512$, the system consists of $N = 262{,}144$ unknowns. A direct factorisation using UMFPACK (a publicly available multifrontal sparse LU solver [24]) serves as the ground-truth reference solve throughout the benchmarking and training phases; its solution defines the exact discrete wavefield against which all approximate solves and learned warm starts are compared.

3.4. Matrix structure and spectral properties

The resulting matrix A_ω is sparse, complex-valued, and non-Hermitian. Its algebraic difficulty is defined by the discrete counterparts of the properties established in Section 2.5.

Throughout this section and the remainder of the thesis, $\|\cdot\|$ applied to a matrix or linear operator denotes the operator norm, also called the induced 2-norm. For a linear map T between normed spaces it is defined as

$$\|T\| \;=\; \sup_{\|v\|=1} \|Tv\|, \tag{3.4}$$

the maximum amplification factor over all unit-norm inputs. When T is a finite matrix this reduces to $\|T\| = \sigma_{\max}(T)$, the largest singular value of T. The operator norm appears both for continuous operators, such as the intertwining defect $\mathcal{E}(\widetilde{\mathcal{T}})$ of equation (2.12), and for discrete matrices, such as the resolvent $(zI - A_\omega)^{-1}$ below; the definition is the same in both cases.

Indefiniteness. Due to the uniform negative diagonal shift $-\omega^2 I$, A_ω is neither positive nor negative definite. Its eigenvalues occupy both sides of the imaginary axis, which invalidates the convergence guarantees of classical stationary iterative methods that rely on diagonal dominance or positive definiteness [25]. As ω increases, the negative mass shift $-\omega^2 I$ moves the discrete Laplacian spectrum further across the origin, so the indefinite part of the operator becomes more pronounced.

Non-normality. The implementation of PML boundary conditions introduces complex-valued off-diagonal entries, breaking self-adjointness. For non-normal matrices, the eigenvalue distribution alone is an insufficient predictor of iterative-solver convergence. The resolvent of A_ω at a point $z \in \mathbb{C}$ is the matrix $(zI - A_\omega)^{-1}$, defined for all z not in the spectrum of A_ω; its operator norm measures how sensitively the solution of the shifted system $(zI - A_\omega)v = g$ responds to perturbations in g. The ε-pseudospectrum of A_ω is defined as

$$\Lambda_\varepsilon(A_\omega) \;=\; \left\{\, z \in \mathbb{C} \,:\, \|(zI - A_\omega)^{-1}\| \,\geq\, \varepsilon^{-1} \,\right\}, \tag{3.5}$$

the set of complex points at which the resolvent norm exceeds ε^{-1}. For self-adjoint matrices, $\Lambda_\varepsilon(A_\omega)$ is an ε-neighbourhood of the spectrum; for non-normal matrices it can extend significantly further, so a favourable eigenvalue distribution is by itself insufficient to guarantee a small GMRES residual at iteration k [8].

Near-zero eigenvalues and conditioning. The discrete Helmholtz operator becomes difficult when eigenvalues of the shifted Laplacian approach the origin. Without a PML, and with a pure Dirichlet boundary closure, the eigenvalues are real and have the form

$$\mu_{mn}(\omega) = \omega^2 - \lambda_{mn}(-\Delta_h), \tag{3.6}$$

up to the sign convention used for A_ω. Hence the matrix becomes nearly singular whenever ω^2 lies close to a discrete Laplacian eigenvalue, and it is singular at exact resonance.

The PML changes this mechanism rather than merely adding damping at the boundary. Complex coordinate stretching makes A_ω non-Hermitian and moves the near-resonant eigenvalues off the real axis. A mode whose real part is close to zero can therefore still have nonzero modulus because the PML contributes an imaginary component. The smallest singular value is then controlled not only by proximity

to a real resonance, but also by how far the PML lifts the corresponding mode into the complex plane. This is why the PML damping schedule affects the observed condition number: adaptive damping can increase the distance of the nearest eigenvalue from the origin, while insufficient damping leaves near-resonant modes almost real and poorly conditioned.

3.5. Grid resolution requirements

To avoid numerical dispersion and maintain physical fidelity, the grid must resolve the wavelength $\lambda = 2\pi/\omega$. The standard heuristic for second-order schemes requires at least ten grid points per wavelength (PPW) [5]:

$$\frac{\lambda}{h} \gtrsim 10. \tag{3.7}$$

In this research, the grid dimension is fixed at $n = 512$ over the unit square, so $h = 1/512 \approx 1.96 \times 10^{-3}$. As shown in Table 3.1, every frequency is well-resolved, ranging from 200 PPW at $\omega = 16$ to 25 PPW at $\omega = 128$. Satisfying local resolution does not eliminate the global *pollution error*, analysed in Section 3.6.

Fixing n rather than h serves two purposes. First, the pollution error grows monotonically with ω across the dataset, so the high-frequency samples represent a strictly more challenging regime for the learned transfer operator: higher frequencies correspond to more oscillatory solutions with finer spatial structure, increasing the task complexity of the mapping $\widetilde{\mathcal{T}} : u_{\omega_L} \mapsto u_{\omega_H}$ and making the approximation error in the intertwining defect $\mathcal{E}(\widetilde{\mathcal{T}})$ larger in operator norm. Second, fixing n keeps the network input dimension constant across frequency pairs: varying n with ω would force the architecture to handle different spatial layouts at different frequencies and conflate spatial-layout effects with frequency physics. This is the same invariance principle that motivates the fixed PML width $n_{\text{pml}} = 112$ in Section 3.7.

Table 3.1: Numerical resolution metrics for the fixed $n = 512$ configuration.

ω	λ	PPW (λ/h)	Resolution status
16	0.393	≈ 200	Over-resolved
32	0.196	≈ 100	Well resolved
64	0.098	≈ 50	Resolved
128	0.049	≈ 25	Resolved

3.6. Dispersion and pollution errors

Satisfying the local sampling condition (3.7) controls the truncation error at every grid point but does not bound the global solution error at high frequency. Two distinct mechanisms are responsible.

Numerical dispersion. A discrete plane wave is the grid function $u_h(\mathbf{x}) = e^{i\mathbf{p}\cdot\mathbf{x}}$, in which $\mathbf{p} = (p_x, p_y)$ is the *numerical wavevector*: the wavevector for which u_h solves the discrete homogeneous operator exactly, in the same way that the physical plane wave with $|\mathbf{k}| = \omega$ solves the continuous one. Its magnitude $|\mathbf{p}|_h$ is the numerical wavenumber and differs from the physical wavenumber ω. Substituting u_h into the five-point Laplacian and using $e^{\pm ip_x h} = \cos(p_x h) \pm i \sin(p_x h)$ collapses the second difference along x to a multiplication factor,

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} = \frac{2\cos(p_x h) - 2}{h^2}\, u_{i,j} = -\frac{4}{h^2} \sin^2\!\left(\tfrac{p_x h}{2}\right) u_{i,j},$$

with the analogous factor along y. Requiring u_h to satisfy the discrete homogeneous operator $-\Delta_h u_h = \omega^2 u_h$ therefore gives the dispersion relation

$$\frac{4}{h^2}\left[\sin^2\!\left(\tfrac{p_x h}{2}\right) + \sin^2\!\left(\tfrac{p_y h}{2}\right)\right] = \omega^2. \tag{3.8}$$

Expanding each term with $\sin^2(\theta) = \theta^2 - \frac{1}{3}\theta^4 + \mathcal{O}(\theta^6)$ at $\theta = p_x h/2$ gives $\frac{4}{h^2}\sin^2(p_x h/2) = p_x^2 - \frac{1}{12}p_x^4 h^2 + \mathcal{O}(p_x^6 h^4)$, so equation (3.8) becomes

$$|\mathbf{p}|^2 \;-\; \frac{h^2}{12}\left(p_x^4 + p_y^4\right) \;+\; \mathcal{O}(h^4) \;=\; \omega^2.$$

To leading order $|\mathbf{p}| = \omega$, so the correction term is $\mathcal{O}(\omega^4 h^2)$ and

$$|\mathbf{p}|_h = \omega\sqrt{1 + \mathcal{O}(\omega^2 h^2)} = \omega + \mathcal{O}(\omega^3 h^2).$$

The discrete phase velocity $\omega/|\mathbf{p}|_h$ therefore lags the physical phase velocity by a relative factor $\mathcal{O}(\omega^2 h^2)$, and the absolute wavenumber error is $\mathcal{O}(\omega^3 h^2)$. Over a domain of fixed diameter the accumulated phase error scales as $\omega^3 h^2$; with h fixed and ω varying it grows as ω^3 rather than vanishing, which is the dispersion contribution to the pollution bound of (3.10) [5, 6].

Pollution effect. The natural measure of solution quality for the Helmholtz problem is the $H^1(\Omega)$ Sobolev norm, defined as

$$\|v\|^2_{H^1(\Omega)} \;=\; \|v\|^2_{L^2(\Omega)} \;+\; \|\nabla v\|^2_{L^2(\Omega)}, \tag{3.9}$$

which measures the approximation error jointly in field amplitude (the L^2 term) and in its first spatial derivatives (the $\|\nabla v\|_{L^2}$ term). For the Helmholtz problem, the spatial gradient is proportional to the acoustic velocity field, so a bound in the H^1 norm simultaneously controls both the pressure and velocity observables.

The inf-sup constant $\beta > 0$ of the discrete operator measures its minimum directional stability:

$$\beta \;=\; \inf_{0 \neq v \in \mathbb{C}^N} \frac{\|A_\omega v\|}{\|v\|_{H^1}}.$$

A small value of β indicates that A_ω is nearly singular in a directional sense: even a modest truncation error in $A_\omega u_h$ is amplified by a factor of β^{-1} upon inversion, producing a large error in u_h itself [6]. Because β decreases as A_ω approaches near-resonant configurations (eigenvalues close to the origin), the amplification grows with ω for a fixed grid.

Combining the dispersion-induced phase error with the inf-sup amplification gives the relative H^1 error bound [6, 5]

$$\frac{\|u - u_h\|_{H^1}}{\|u\|_{H^1}} \;\leq\; C_1\,\omega h \;+\; C_2\,\omega^3 h^2, \tag{3.10}$$

with constants C_1, C_2 independent of ω and h. The first term is the best-approximation (interpolation) error of the resolved wavefield in the H^1 norm, which scales as ωh and is controlled by (3.7); it is distinct from the pointwise $\mathcal{O}(h^2)$ truncation error of the five-point stencil (Section 3.2), the difference being that the relevant quantity here is the solution error in the energy norm rather than the residual of the difference equation. The second term is the pollution contribution; it cannot be removed by satisfying the PPW heuristic pointwise because it grows with the total number of wavelengths in the domain. For the configuration of this thesis h is fixed and ω varies, so the pollution term grows as $\omega^3 h^2$ across the four operating frequencies.

The growth of the pollution term across the four operating frequencies is the mechanism described in Section 3.5; higher-order discretisations that reduce it are recorded as future work in Section 9.2.

3.7. Absorbing boundary conditions

The continuous PML construction of Section 2.4 is discretised on the same uniform Cartesian grid as the interior operator. The coordinate-stretching mechanism defined there is applied with discrete parameters n_{pml} and σ_0 specific to the grid: a fixed layer width of $n_{\text{pml}} = 112$ grid points is used across all frequencies, and the peak damping σ_0 is calibrated per frequency as described below.

This fixed-width design is a critical architectural choice: it ensures the physical interior domain remains a constant 288×288 grid for all samples. If n_{pml} were frequency-adaptive, the neural network would be forced to handle varying spatial layouts, conflating boundary geometry with frequency physics. The reflection coefficient R is defined as the ratio of the amplitude of a normally-incident plane wave reflected back into the physical interior to the amplitude of the same wave incident on the PML interface, in the limit of vanishing finite-difference truncation error. The peak absorption σ_0 is calibrated per frequency to keep the reflection coefficient at $R \leq 10^{-2}$ across all frequencies, according to the schedule in Table 3.2.

The numerical values of σ_0 in Table 3.2 are the outcome of a deliberate balance between boundary absorption and algebraic stability. Increasing σ_0 reduces the boundary reflection R, while the ratio σ_0/ω

controls how strongly the PML shifts near-resonant modes away from the real axis. This shift is important because, without sufficient PML damping, modes whose real eigenvalues lie close to zero remain nearly singular and dominate the condition number of A_ω. The schedule selected here therefore uses stronger damping at higher frequencies to lift these modes farther into the complex plane, while avoiding unnecessarily large damping that would increase non-normality without improving the solver-facing metrics. The full justification of the schedule follows the table.

Throughout this thesis the PML width is fixed at $n_{\text{pml}} = 112$ grid cells on each side of the domain. The total computational grid is the unit square at $n \times n$ with $n = 512$ and $h = 1/512$; the physical interior therefore occupies the central $(n - 2n_{\text{pml}}) \times (n - 2n_{\text{pml}}) = 288 \times 288$ sub-grid, independent of frequency.

Table 3.2: PML absorption schedule. A sub-linear scaling at high frequencies balances boundary absorption against matrix conditioning.

ω (rad/s)	σ_0	Scaling regime
16	42.5	Linear
32	85.0	Linear
64	120.0	Sub-linear
128	180.0	Sub-linear

Derivation of the reflection bound. For the quadratic damping profile $\sigma(x) = \sigma_0(x/\delta)^2$ with layer thickness $\delta = n_{\text{pml}}\, h$, the total damping accumulated by a wave traversing the layer once is

$$\int_0^\delta \sigma(x)\, \mathrm{d}x \;=\; \frac{\sigma_0\, \delta}{3} \;=\; \frac{\sigma_0\, n_{\text{pml}}\, h}{3}.$$

The reflection coefficient accounts for both the outgoing and the returning wave paths (two traversals of the layer), giving a round-trip damping exponent of $2\sigma_0 n_{\text{pml}} h/3$ and therefore

$$R \;=\; \exp\!\big(-\tfrac{2}{3}\,\sigma_0\, n_{\text{pml}}\, h\big)\,. \tag{3.11}$$

Figure 3.2 shows the quadratic damping profile at the four operating frequencies: σ is identically zero at the interior–PML interface, ramps quadratically across the 112 grid cells of the absorbing layer, and reaches its peak σ_0 at the outer wall.

Rationale for the schedule. The values of σ_0 in Table 3.2 are not chosen by the reflection bound alone. Two competing constraints determine the schedule.

The first constraint is the reflection target. From equation (3.11), $R \le 10^{-2}$ requires $\sigma_0 \ge (3 \ln 100)/(2\, n_{\text{pml}}\, h)$, a fixed lower bound that depends only on the grid. A purely linear schedule $\sigma_0 \propto \omega$ satisfies this constraint at all four operating frequencies.

The second constraint is solver performance. The stretching factor $s(x) = 1 + i\sigma(x)/\omega$ introduces imaginary off-diagonal and diagonal entries into A_ω whose magnitude scales with σ_0/ω; larger values of σ_0 therefore increase the non-normality of the matrix (Section 3.4), widening the pseudospectrum and requiring FGMRES to handle a harder operator independently of the boundary reflection. At $\omega = 64$ and $\omega = 128$, the linear schedule yields σ_0 values that exceed the minimum needed for $R \le 10^{-2}$ and produce measurably higher FGMRES iteration counts than necessary. The sub-linear values in Table 3.2 are the smallest σ_0 at each of those frequencies that simultaneously achieves $R \le 10^{-2}$ and minimises the FGMRES iteration count on the corresponding unpreconditioned system; they were selected by sweeping σ_0 at fixed ω and reading off the minimum of the combined cost. At $\omega = 16$ and $\omega = 32$, the linear schedule happens to coincide with this minimum, so no correction is applied.

The smooth onset of the quadratic profile eliminates the impedance mismatch that a step profile would introduce. The algebraic role of the PML is subtler than a pure conditioning penalty: the stretching factor contains the ratio $\sigma(x)/\omega$, so near-resonant modes that would remain close to the origin for a Dirichlet closure acquire an imaginary component whose size is governed by the amount of damping seen by that mode. Increasing σ_0/ω therefore tends to lift these modes away from the origin and can increase

the smallest singular value of A_ω. This is the mechanism behind the separation between constant and adaptive damping in the condition number diagnostic: with adaptive damping, the eigenvalue closest to the origin is pushed farther into the complex plane.

This trade-off is fundamental to PML design for high-frequency problems: tighter absorption increases the imaginary shift in A_ω, which helps physically but worsens the algebraic conditioning that the Krylov solver must overcome. The selected schedule in Table 3.2 maintains acceptable boundary absorption while ensuring the high-frequency systems are not dominated by PML-induced near-singularities.

The 1D profile of Figure 3.2 is applied along each spatial axis independently, producing the two-dimensional absorbing frame shown in Figure 3.3. The 112-cell collar surrounds the 288×288 physical interior on all four sides; the corner regions, where σ acts in both x and y, provide additional damping for waves propagating diagonally toward the outer boundary.

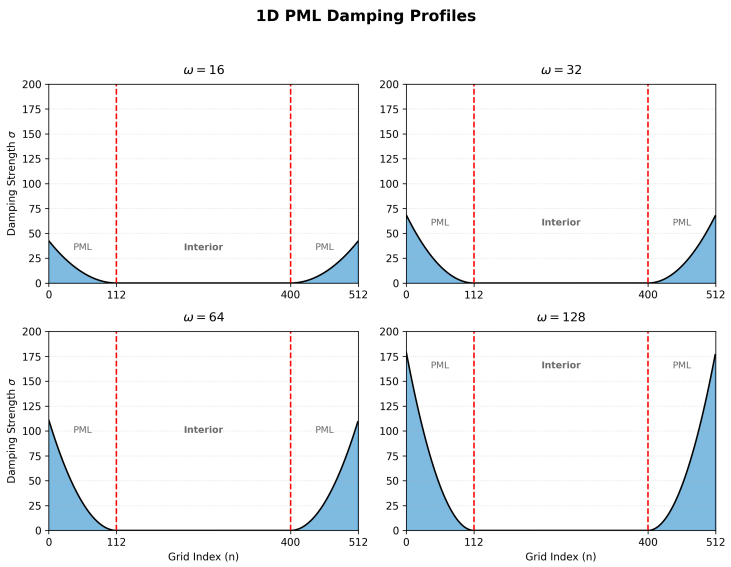


Figure 3.2: Quadratic damping profile $\sigma(x)$ across the 112-cell absorbing layer for the four operating frequencies. The horizontal axis is the grid-cell index measured outward from the physical-domain/PML interface. The vertical axis gives the damping strength σ in the same angular-frequency units as ω (rad/s). The profile vanishes at the interior boundary and reaches its peak value σ_0 at the outer wall; the corresponding σ_0 values are listed in Table 3.2.

Modified stencil inside the PML. The five-point stencil (3.2) is modified inside the absorbing layer through the coordinate-stretching factor $s(x) = 1 + i\sigma(x)/\omega$. Because $\sigma(x)$ varies from grid point to grid point inside the PML, the stretching factor must be evaluated at the midpoint between two adjacent nodes rather than at either node alone. The face-averaged value $s_{i+1/2,\,j} = \frac{1}{2}(s_{i,j} + s_{i+1,j})$ approximates s at the midpoint of the cell face between grid points (i, j) and $(i + 1, j)$; this is the natural location because the finite-difference quotient $(u_{i+1,j} - u_{i,j})/h$ approximates $\partial_x u$ there, so the stretching factor must be co-located. Arithmetic averaging maintains second-order accuracy without introducing additional stencil points.

For a grid point (i, j) within the PML, the second derivative in the x-direction is therefore approximated as [22]

$$\left[\partial_{xx} u\right]_{i,j} \;\approx\; \frac{1}{s_{i,j}\, h^2}\left[\frac{u_{i+1,j} - u_{i,j}}{s_{i+1/2,j}} \;-\; \frac{u_{i,j} - u_{i-1,j}}{s_{i-1/2,j}}\right], \tag{3.12}$$

with the analogous expression in the y-direction. Because $s(x) \in \mathbb{C}$ inside the PML, equation (3.12) introduces complex-valued entries into both the off-diagonal and the diagonal of A_ω at every grid point in the absorbing region: the off-diagonals through the face-averaged factors $s_{i\pm1/2,j}$, and the diagonal through the prefactor $1/(s_{i,j}h^2)$ together with the mass term $-\omega^2$. The five-point sparsity pattern is preserved; only the values of the non-zero entries change. This is the algebraic origin of the non-Hermitian structure analysed in Section 3.4.

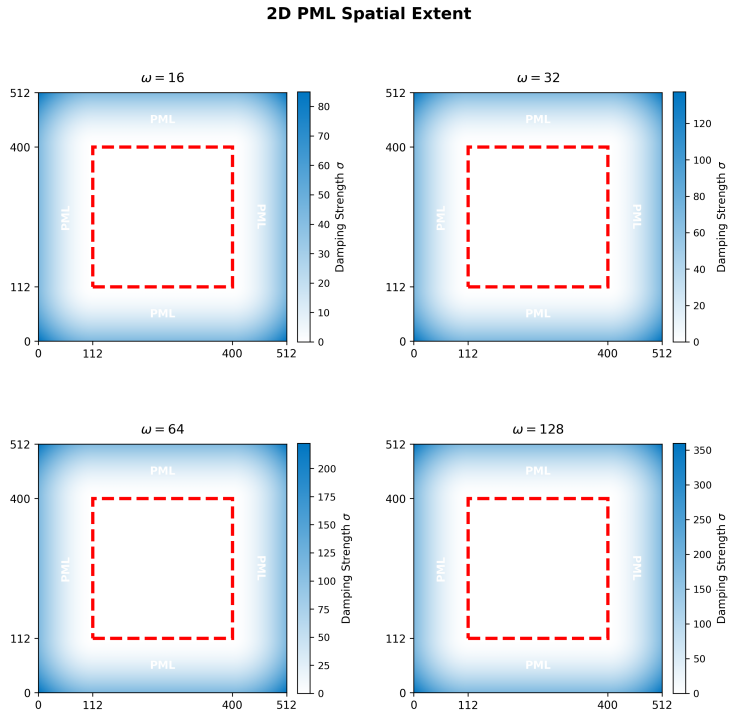


Figure 3.3: Two-dimensional PML footprint. The fixed 112-point absorbing collar (shaded) surrounds the 288×288 physical interior (dashed) on all four sides at every operating frequency, keeping the spatial layout independent of ω.

4

Classical solution methods

This chapter reviews established numerical strategies for solving the discretised Helmholtz system

$$A_\omega u_\omega = b_\omega \tag{4.1}$$

derived in Chapter 3. The purpose is not to give a complete survey of Helmholtz solvers, but to identify the structural obstructions that make the high-frequency problem expensive and to define the classical baseline against which the learned component of this thesis must be judged.

A key starting point is that A_ω is sparse. The five-point finite-difference stencil produces at most five nonzero entries per row, so storage and matrix–vector products scale as $\mathcal{O}(N)$ in the number of unknowns. The difficulty is therefore not matrix density, but spectral structure. As ω increases, the discrete operator becomes increasingly indefinite and, after the introduction of a PML, non-normal. Its spectrum and pseudospectrum move in ways that degrade Krylov convergence, preconditioning, and coarse-grid correction.

The chapter also makes a critical distinction that is important for the rest of the thesis. Learned frequency transfer should not be compared only with weak classical strawmen. In favourable linear settings, classical methods can already exploit Green's functions, superposition, low-rank structure, and frequency continuation. The learned method developed later is therefore not presented as a replacement for the best classical linear Helmholtz solvers. Instead, the linear FD/PML problem is used as a controlled setting in which to test whether a learned field prediction, once inserted as a warm start, accelerates the high-frequency CSL-preconditioned FGMRES solve.

4.1. Krylov subspace methods

This section establishes why high-frequency Helmholtz problems require iterative methods, which iterative method is used in this thesis, and what its convergence depends on.

Despite the sparsity of A_ω, direct factorisation remains costly. With a nested-dissection reordering [26], sparse LU factorisation in two dimensions requires $\mathcal{O}(N^{3/2})$ operations and produces factors with $\mathcal{O}(N\log N)$ nonzeros; each back-solve then costs $\mathcal{O}(N\log N)$. In practice, multifrontal sparse direct solvers such as UMFPACK realise these advantages over dense factorisation and are used in this thesis for reference solves and data generation.

Two scaling regimes must be distinguished. If the grid is refined to keep a fixed number of points per wavelength, then $N = \mathcal{O}(\omega^2)$ in two dimensions. The sparse factorisation cost therefore scales as $\mathcal{O}(\omega^3)$. If the grid is refined further to suppress pollution error, the constraint is stronger. For second-order finite differences one obtains $h = \mathcal{O}(\omega^{-3/2})$, hence $N = \mathcal{O}(\omega^3)$ and a factorisation cost of $\mathcal{O}(\omega^{9/2})$. Both estimates assume grid refinement under the pollution constraint; the experiments of this thesis instead hold the grid fixed, so neither growth is paid, and the sparse five-point structure keeps every reference solve far below the dense $\mathcal{O}(N^3)$ cost.

The experiments in this thesis follow a third regime. The grid size is held fixed across the frequency pairs, so the number of unknowns does not grow with ω. What changes is the spectral structure of A_ω. This fixed-grid regime isolates the solver-facing question: even when the matrix size is fixed, do the

iteration counts and residual histories deteriorate with frequency, and can a learned warm start reduce that deterioration?

Krylov subspace methods approximate the solution in the nested spaces

$$\mathcal{K}_m(A_\omega, r_0) = \text{span}\left\{r_0, A_\omega r_0, A_\omega^2 r_0, \ldots, A_\omega^{m-1} r_0\right\}, \tag{4.2}$$

where $r_0 = b_\omega - A_\omega u_0$ is the initial residual. The approximate solution $u_m \in u_0 + \mathcal{K}_m$ is then chosen by minimising an appropriate residual or error measure over this subspace.

The Generalised Minimal Residual method, GMRES, is the standard Krylov method for non-Hermitian systems [27]. The Arnoldi process constructs an orthonormal basis for $\mathcal{K}_m$, and GMRES selects the iterate that minimises $\|b_\omega - A_\omega u_m\|_2$ over the affine Krylov space. This gives monotone non-increase of the true residual norm for unrestarted GMRES in exact arithmetic, and the exact solution is reached after at most N iterations. In practice, the relevant quantity is not this worst-case finite termination result, but the number of iterations needed to reach a prescribed tolerance.

Storing the full Arnoldi basis requires $\mathcal{O}(mN)$ memory, where m is the number of Krylov iterations. Restarted GMRES reduces storage but may stagnate. Flexible GMRES, FGMRES, stores both the Krylov basis and the preconditioned search directions, allowing the preconditioner to vary from one iteration to the next [13]. This flexibility becomes important if a learned or adaptive preconditioner is used later, since such a preconditioner may depend on the current residual.

For a normal matrix, GMRES convergence can be understood largely from the eigenvalue distribution. The Helmholtz matrix with PML is non-normal, so eigenvalues alone are not sufficient. The relevant object is the ε-pseudospectrum $\Lambda_\varepsilon(A_\omega)$, defined in equation (3.5), which can extend far beyond the eigenvalues for highly non-normal matrices [8].

For high-frequency Helmholtz systems, the dimension of the Krylov space required to reach a fixed tolerance typically grows with frequency unless an effective preconditioner is used [9]. This is the first reason a learned initial guess, by itself, cannot be expected to solve the whole problem: it changes r_0, but it does not change the operator that governs the subsequent Krylov convergence.

4.2. Preconditioning strategies

Preconditioning replaces the original system by an equivalent one whose coefficient matrix is more favourable for Krylov iteration. In right preconditioning one solves

$$A_\omega M^{-1} y = b_\omega \qquad u = M^{-1} y, \tag{4.3}$$

where $M \approx A_\omega$ and the action $v \mapsto M^{-1} v$ can be computed at much lower cost than an exact solve with A_ω. The aim is to make $A_\omega M^{-1}$ closer to the identity, or at least to cluster its spectrum and pseudospectrum in a region where GMRES polynomials can reduce the residual rapidly.

Classical elliptic preconditioners such as Jacobi scaling, Gauss–Seidel, and incomplete LU factorisation are designed for positive-definite or diagonally dominant operators. The Helmholtz operator violates these assumptions. Sparse Cholesky factorisation is excluded because A_ω is not Hermitian positive definite. ILU-type methods are formally applicable, but their effectiveness deteriorates as the operator becomes more indefinite and non-normal.

The difficulty is structural. As ω increases, the shift $-\omega^2 I$ moves eigenvalues of the Laplacian through the origin and onto both sides of the complex plane. Near-zero eigenvalues and a wide pseudospectrum make the inverse hard to approximate uniformly. A preconditioner that works at one frequency is therefore not automatically effective at a higher frequency. This motivates preconditioners designed specifically for Helmholtz problems, of which the Complex Shifted Laplacian is the most important for this thesis.

4.3. The Complex Shifted Laplacian preconditioner

The Complex Shifted Laplacian, CSL, preconditioner is the standard classical remedy for the indefiniteness of the Helmholtz operator [9]. It introduces artificial damping by replacing the original operator with

$$M_{\text{CSL}} = -\Delta_h - (1 + i\beta)\omega^2 I. \tag{4.4}$$

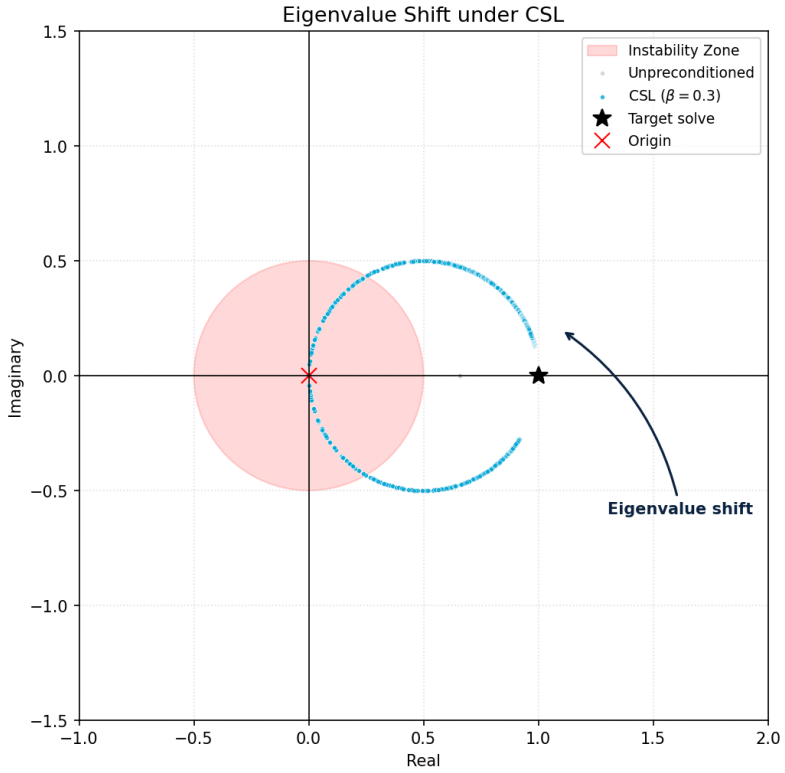


Figure 4.1: Eigenvalues of $A_\omega M_{\text{CSL}}^{-1}$ in the complex plane at $\omega = 64$ with $\beta = 0.3$. The complex shift moves the indefinite Helmholtz spectrum away from the origin and towards the FGMRES target $z = 1$, but the eigenvalues do not collapse to a single point. The remaining near-origin arc is the part of the preconditioned problem where a warm start can still earn its place in Chapter 7: CSL improves the operator, but it does not make the high-frequency solve frequency-independent, so room remains for a better initial iterate to reduce the iteration count.

where $\beta > 0$ controls the strength of the complex shift. The imaginary part moves the near-singular part of the spectrum away from the origin and makes the shifted system easier to invert approximately.

There is a trade-off. A larger β makes M_{CSL} easier to invert, but also makes it a worse approximation of A_ω. A smaller β gives a more faithful preconditioner, but the shifted system becomes harder to solve. Values such as $\beta \in [0.3, 0.5]$ are commonly used as practical compromises [11].

Figure 4.1 illustrates the effect of CSL. The preconditioner succeeds in its primary task: the most dangerous part of the spectrum is moved away from the origin. A spectral analysis of the CSL-preconditioned operator shows that the eigenvalues cluster on a circle in the complex plane whose geometry fixes the optimal shift and explains the mesh dependence of GMRES convergence [28]. However, the preconditioned spectrum still forms an extended arc rather than a tight cluster at $z = 1$. The remaining spread is what the outer FGMRES iteration must resolve. As the frequency increases, this preconditioned problem remains harder, so CSL improves the solver but does not give frequency-independent convergence.

In the experiments of Chapter 7 the value $\beta = 0.3$ is used. This value is chosen because it keeps the shifted operator close enough to the original Helmholtz operator while still providing sufficient complex damping to make the CSL solve stable. A larger value such as $\beta = 0.5$ would generally make the shifted system easier to invert, but at the cost of a less faithful approximation to A_ω and therefore a less informative outer FGMRES baseline. The choice $\beta = 0.3$ is thus the more demanding and more solver-facing compromise: it tests learned warm starts against a CSL preconditioner that is strong, but not so heavily damped that the original Helmholtz residual geometry is washed out.

In large-scale computations, the action $M_{\text{CSL}}^{-1} v$ is usually approximated by one or more multigrid cycles applied to the shifted operator [29]. The resulting algorithm has two levels of iteration: an outer FGMRES

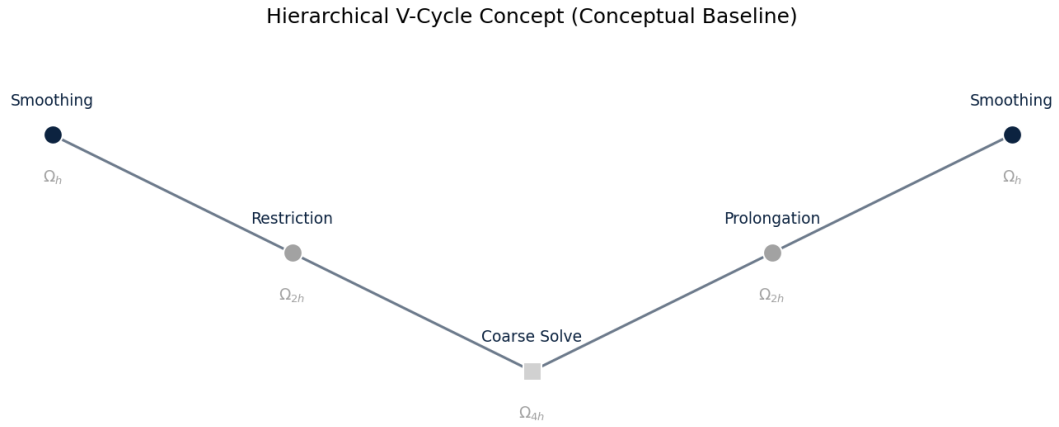


Figure 4.2: Schematic of a classical geometric multigrid V-cycle on a three-level hierarchy. The residual is restricted to coarser grids, corrected there, and prolongated back to the fine grid. The same restriction –> coarse solve –> prolongation pattern motivates the learned residual-transfer perspective developed later.

iteration for the original Helmholtz system, and an inner approximate solve with the shifted Laplacian. This CSL–FGMRES combination is the main classical solver used in the experiments of this thesis.

4.4. One iteration of CSL-preconditioned FGMRES

It is useful to make the place where the preconditioner enters the outer iteration explicit, because the warm-start analysis of later chapters depends on it. Given the current iterate x_k and residual $r_k = b - A_{\omega_H} x_k$, the preconditioned step applies M_{CSL}^{-1} to r_k, expands the Krylov subspace with the resulting vector, and forms the next iterate by minimising the residual over the updated subspace:

$$z_k = M_{\text{CSL}}^{-1} r_k, \qquad \mathcal{K}_{k+1} = \text{span}\{z_0, A_{\omega_H} z_0, \ldots, A_{\omega_H}^k z_0\}, \qquad x_{k+1} = \arg\min_{x \in x_0 + \mathcal{K}_{k+1}} \|b - A_{\omega_H} x\|_2. \quad (4.5)$$

The preconditioner acts on residuals; the minimisation is performed against the original operator A_{ω_H}. This is the structure that Chapter 5 reuses when a learned solution transfer supplies the initial guess x_0 at $k = 0$, while the CSL preconditioner acts at every subsequent iteration.

4.5. Multilevel and multigrid

Multigrid methods are optimal for many elliptic problems because they separate error components by scale. A smoother damps high-frequency error on the current grid, while a coarse-grid correction removes error components that appear smooth on that grid. These operations are combined in a V-cycle: smooth, restrict the residual to a coarser grid, solve or approximately solve the coarse problem, prolongate the correction, and smooth again.

Standard multigrid fails for the Helmholtz operator for two coupled reasons. First, the operator is indefinite. Classical smoothers such as Gauss–Seidel are designed to damp oscillatory error components for elliptic operators, but Helmholtz supports oscillatory modes that may not be damped. For certain frequencies and grids, the smoother can even amplify relevant error modes [9].

Second, the coarse-grid equation may no longer represent the same wave problem. Coarse-grid correction is meaningful only if the wave remains resolved on the coarse grid. Once the fine grid is already close to its resolution limit, doubling the grid spacing can violate the points-per-wavelength requirement. The correction computed on the coarse grid then carries the wrong phase and no longer improves the fine-grid solution reliably [23].

Several modifications have been proposed. Shifted multigrid applies multigrid to the complex-shifted operator used in CSL. Algebraic multigrid constructs coarse spaces from the matrix rather than from an

explicit grid [30]. Sweeping preconditioners exploit directional structure in wave propagation and can achieve near-linear complexity in favourable two-dimensional settings [31]. These methods are highly important competitors. However, none removes all frequency dependence in the general high-frequency Helmholtz problem, and each relies on structure that may be unavailable or less useful in more complicated settings.

This observation is important for the learned approach. The learned method should not be motivated by claiming that classical solvers ignore all structure. They do not. Rather, the motivation is that there remains a gap between strong structure-exploiting classical solvers and the ideal goal of frequency-independent, many-query, high-frequency wave solves.

4.6. Classical reuse of structure across sources and frequencies

The previous sections focused on Krylov iteration and preconditioning. A second class of classical methods attacks the problem by reusing linear structure across sources, frequencies, or parameter values. These methods are especially important because they exploit some of the same coherence that motivates learned frequency transfer.

For a homogeneous linear medium, the solution can be represented using the Green's function. If the source is smooth or effectively low-dimensional, it can be expanded in a small quadrature or spectral basis. One can then solve for a modest number of basis sources and represent new solutions as linear combinations of precomputed responses. In far-field regimes, off-diagonal blocks of the Green's function are often low-rank, so randomized probing or compressed representations can approximate the solution operator using far fewer solves than a dense sampling of all possible right-hand sides.

Frequency-sweep methods exploit nearby frequencies in a different way. Subspace recycling, rational approximants, and Padé-type continuation methods reuse information from previous linear systems to accelerate later ones. These approaches are natural competitors to learned frequency transfer because they also recognise that consecutive Helmholtz solves are not independent.

The existence of these methods changes the interpretation of the present thesis. In the favourable linear regime, a neural network is not automatically the most efficient way to exploit structure. Linearity and superposition give classical methods a major advantage: they can use known physics directly rather than learning it from examples. The learned method is therefore most credible in regimes where these classical shortcuts weaken, for example for non-smooth or spatially distributed sources, near-source quantities, many interacting parameters, or nonlinear wave problems for which no Green's-function superposition principle is available.

The experiments in this thesis deliberately remain in a linear FD/PML setting. This is not because this is the regime in which learning is expected to have its largest final advantage. It is because the linear setting provides a controlled diagnostic. If a learned frequency-transfer map cannot accelerate the linear solve, where the structure is most favourable and most fully understood, it cannot be expected to work reliably in harder nonlinear or high-parameter regimes. The linear problem is therefore used as a baseline of difficulty rather than as the final deployment claim.

4.7. Breakdown at high frequency

The combined spectral and algebraic properties above lead to a persistent growth of computational cost as ω increases. Three mechanisms are most important.

The first mechanism is spectral drift. As ω increases, the shift $-\omega^2 I$ moves the spectrum of the discretised operator through and around the origin. A fixed preconditioner can cluster the spectrum only if it remains a good approximation of the inverse across the relevant spectral region. For Helmholtz, that region changes with frequency, so a preconditioner designed for one frequency degrades at another.

The second mechanism is non-normality. In the presence of a PML, the coordinate stretching makes the matrix non-normal. Even if the eigenvalues of the preconditioned operator look acceptable, the pseudospectrum may remain large. FGMRES convergence can then be slow because the residual polynomial must be small on a much larger effective set than the eigenvalues alone suggest [8].

The third mechanism does not operate in the experiments of this thesis but is recorded because it governs realistic deployment: grid growth. In practical high-frequency deployment, the grid must usually

be refined as the wavelength decreases. Keeping a fixed number of points per wavelength already gives $N = \mathcal{O}(\omega^2)$ in two dimensions. Suppressing pollution error requires still finer grids. The experiments of this thesis hold the grid fixed, so this growth is not paid directly, but it remains the background reason why high-frequency Helmholtz solves become increasingly expensive in realistic simulations.

These mechanisms imply that CSL-preconditioned FGMRES is expected to be the strongest classical reference method in the experiments, but not a frequency-independent one. A learned warm start may reduce the number of iterations to a fixed tolerance by starting the iteration from a better iterate, but it should not be expected to change the asymptotic convergence slope, because it does not change the spectrum or pseudospectrum of $A_{\omega_H} M_{\mathsf{CSL}}^{-1}$. Its effect is therefore confined to the finite-budget part of the convergence history, and is not reliably measured by the initial residual it produces.

Expected consequences for the experiments. The preceding analysis gives three expectations for the experimental chapters. First, GMRES without preconditioning, and GMRES with classical elliptic preconditioners such as Jacobi, Gauss–Seidel, and incomplete LU, should deteriorate quickly as the operator becomes more indefinite and non-normal, since none of these is designed for an indefinite, non-normal operator.

Second, CSL-preconditioned FGMRES should be the strongest classical baseline, but its iteration count should still grow with frequency.

Third, a low-frequency solution used only as an initial guess does not modify the high-frequency operator, the CSL preconditioner, or the Krylov convergence mechanism. It changes only the initial iterate, so its effect is confined to the finite-budget part of the convergence history rather than the asymptotic slope. That effect is not adequately measured by the initial residual alone. The warm start can remove error components that the iteration would otherwise have to rebuild, and those components do not reappear, so it can lower the final iteration count even when the initial residual it produces is no smaller, or is larger, than the zero-start residual. The reverse is also possible: a warm start that lowers the initial residual on already-fast directions may save no iterations at all. Initial residual and iteration count are therefore distinct measures, and the second is the binding one.

These expectations define the validation standard for the learned component. A learned frequency-transfer map is not successful merely because its predicted field resembles u_{ω_H}. For solver purposes the relevant question is whether it whether it reduces the number of CSL-preconditioned FGMRES iterations required to reach a prescribed tolerance; the iteration count is the binding measure used in this thesis, for the cost-accounting reasons of Section 6.6.2. The initial relative residual

$$\frac{\|b_{\omega_H} - A_{\omega_H} u_0\|_2}{\|b_{\omega_H}\|_2} \tag{4.6}$$

and the finite-budget FGMRES residual after a fixed number of steps are informative intermediate quantities, but neither is the criterion on its own. A learned warm start that improves field-error metrics while leaving the initial residual elevated may still reduce the iteration count, and the classical analysis of this chapter does not rule that out; what it does rule out is any change to the asymptotic convergence slope, which the warm start cannot affect because it does not alter the spectrum or pseudospectrum of $A_{\omega_H} M_{\mathsf{CSL}}^{-1}$.

4.8. Open problem and motivation for learned residual transfer

The open problem addressed by this thesis is not that all classical Helmholtz methods ignore structure. Many do exploit structure, and some do so very effectively in favourable linear regimes. The open problem is narrower and more solver-facing: can information from a lower-frequency wavefield be inserted into a high-frequency Krylov solve in a way that measurably accelerates it?

In full-waveform inversion and related frequency-sweep workflows, lower frequencies are solved before higher frequencies. Thus a low-frequency wavefield u_{ω_L} is available by the time the high-frequency system

$$A_{\omega_H} u = b_{\omega_H} \tag{4.7}$$

is assembled. Helmholtz solutions at related frequencies share propagation geometry: their amplitude envelopes and phases are linked by the wave paths and the frequency gap [12]. This coherence motivates a transfer map from ω_L to ω_H.

The simplest use of such a map is a warm start. A network predicts a high-frequency field, and this field is used as the initial guess u_0 for CSL-preconditioned FGMRES. This can lower the iteration count by starting from a better iterate. However, a warm start acts only once. It cannot change the Krylov subspace generated by the preconditioned high-frequency operator after the initial residual has been formed. Therefore, even an accurate solution-transfer map is structurally limited as a solver component.

A stronger approach is learned residual transfer. In that setting, the learned component acts inside the iteration. The high-frequency residual is restricted or transferred to a lower-frequency representation, a low-frequency correction is computed, and the result is prolongated back to the high-frequency space. This is analogous in structure to a multigrid V-cycle, but with frequency rather than grid resolution defining the hierarchy. Because the learned preconditioner may vary with the current residual and iteration state, FGMRES is the natural outer Krylov method.

This distinction motivates the next chapter. Chapter 5 derives the frequency-transfer operator, separates solution transfer from residual and correction transfer, and shows why training only on field error is not enough. The central analytical question becomes: when does the information in a lower-frequency solution still accelerate the high-frequency solve once it has been inserted as a warm start and the CSL-preconditioned iteration has run? The residual-amplification mechanism derived in the next chapter explains why this question cannot be settled by the initial residual alone, and it is the thread that the rest of the thesis follows.

Part II

Method

5

The frequency transfer operator

The preceding chapters introduced the Helmholtz equation, its finite difference discretisation, absorbing boundary treatment, and the classical iterative methods used later in the thesis. A recurring observation is that solving a high-frequency Helmholtz system is much harder than solving a low-frequency one on the same medium, even when the underlying propagation geometry is identical. Classical solvers approach each frequency in isolation and do not exploit any relationship between solutions at neighbouring frequencies.

This chapter introduces frequency transfer as a way to use that relationship. In a homogeneous medium, changing the frequency changes the wavelength and amplitude scale, but not the underlying propagation geometry. The low-frequency wavefield therefore already contains coherent information about where energy travels and how it decays. The mathematical object that captures this relationship is the frequency transfer operator: a map that takes a solution at one frequency as its input and returns the solution at a higher frequency as its output. Section 5.1 formalises the physical intuition behind this object, and Section 5.2 defines it precisely as a map between the low- and high-frequency solution manifolds, that is, the sets of fields that exactly satisfy the respective Helmholtz systems for some source.

The remainder of the chapter then asks how such an operator can be used. The first and most direct application is as a warm start for a high-frequency solve, treated in Section 5.3. The same building block also suggests a multilevel preconditioning structure, which Section 5.4 develops by separating three transfer objects that share frequency-transfer intuition but play distinct algebraic roles. Sections 5.5–5.6 then introduce a refinement that becomes important once the operator is placed inside a Krylov method: small field errors can correspond to non-negligible residuals because of how the discrete operator amplifies certain components. This refinement motivates the learning objectives and the model requirements summarised in Sections 5.7–5.8.

A central distinction in this chapter is the difference between a *warm start* and a *preconditioner*. A warm start changes only the initial guess x_0 before the Krylov iteration begins. After that point, the Krylov method and the CSL preconditioner proceed exactly as usual. A preconditioner, by contrast, is applied inside every Krylov iteration and changes the search directions generated from the current residuals. This thesis therefore treats learned frequency transfer first as a solution-transfer warm start, and only later discusses what additional structure would be required to turn the same idea into a genuine learned preconditioner.

5.1. Frequency coherence of Helmholtz fields

Consider the time-harmonic Helmholtz equation in a homogeneous medium with unit wave speed,

$$\mathcal{L}_\omega u := \left(-\Delta - \omega^2 I\right) u = f, \tag{5.1}$$

where ω is the angular frequency. In two-dimensional free space, the outgoing Green's function for a point source at $\mathbf{x}_s$ is [21]

$$G_\omega(\mathbf{x}, \mathbf{x}_s) = \frac{i}{4} H_0^{(1)}(\omega r), \qquad r = \|\mathbf{x} - \mathbf{x}_s\|_2, \tag{5.2}$$

The function $H_0^{(1)}$ is the zeroth-order Hankel function of the first kind, defined as the complex combination $H_0^{(1)} = J_0 + iY_0$ of the Bessel functions of the first and second kind. It is the outgoing fundamental

solution of the two-dimensional Helmholtz operator: J_0 supplies the standing-wave part that stays finite at the source, Y_0 supplies the logarithmic singularity carried by a point source, and the combination $J_0 + iY_0$ selects the radially outgoing wave that satisfies the Sommerfeld radiation condition. Its behaviour away from the source, which is the part that controls the propagating field, follows from the large-argument asymptotic value

$$H_0^{(1)}(z) \sim \sqrt{\frac{2}{\pi z}}\,\exp(i(z-\pi/4))\,, \qquad |z| \to \infty. \tag{5.3}$$

Substituting (5.3) into (5.2) gives the far-field form

$$G_\omega(\mathbf{x}, \mathbf{x}_s) \sim C_\omega r^{-1/2} e^{i\omega r}, \qquad C_\omega = \frac{i}{4}\sqrt{\frac{2}{\pi\omega}}e^{-i\pi/4}. \tag{5.4}$$

Equation (5.4) separates the field into three components. The factor $r^{-1/2}$ is the two-dimensional geometric spreading envelope. It depends on propagation distance, but not on the frequency. The constant C_ω contains the global frequency-dependent amplitude scale, with $|C_\omega| \propto \omega^{-1/2}$. The phase factor $e^{i\omega r}$ contains the oscillation rate. Increasing ω therefore increases the number of phase cycles along the same ray path while preserving the large-scale propagation geometry.

For a doubled-frequency pair $\omega_H = 2\omega_L$, the far-field amplitude ratio is

$$\frac{|C_{\omega_H}|}{|C_{\omega_L}|} = \sqrt{\frac{\omega_L}{\omega_H}} = \frac{1}{\sqrt{2}}. \tag{5.5}$$

This ratio should not be overinterpreted. A high-frequency field is not a pointwise rescaling of the low-frequency field, because the phase changes along every propagation path. The point is weaker and more useful: the two solutions are coherent. They share geometry, and they differ in phase, amplitude scale, and boundary-layer behaviour in ways that are themselves structured by the frequency change.

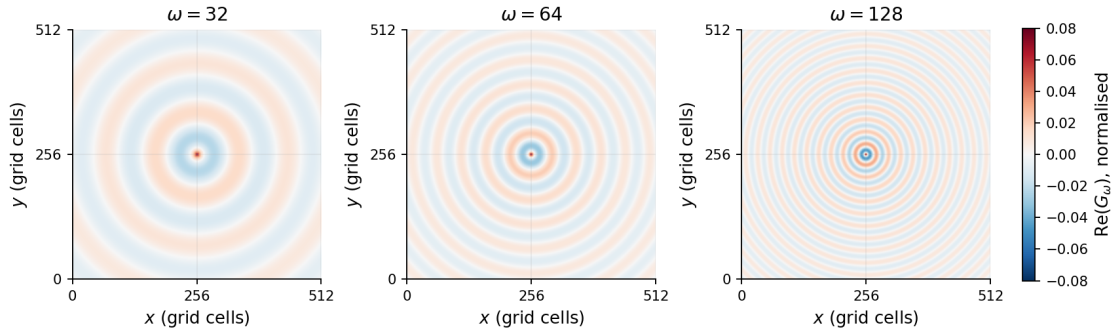


Figure 5.1: Frequency coherence of homogeneous Helmholtz fields. The real part of the free-space Green's function is shown for several frequencies with one centred source. The propagation geometry is common to all panels, while the oscillation rate increases with frequency. Each panel is normalised by its maximum absolute value so that the visual comparison emphasises phase structure rather than absolute amplitude.

This physical coherence is the reason a data-driven transfer operator is a reasonable object to learn. The low-frequency field is a compressed description of propagation geometry. The high-frequency field asks for the same geometry at a shorter wavelength, with a quantitatively predictable shift in amplitude scale. A learned operator must capture both the shared structure and the frequency-dependent transformation between the two.

5.2. Exact solution transfer and the intertwining defect

Let u_{ω_L} and u_{ω_H} solve two Helmholtz problems with the same source,

$$\mathcal{L}_{\omega_L} u_{\omega_L} = f, \qquad \mathcal{L}_{\omega_H} u_{\omega_H} = f. \tag{5.6}$$

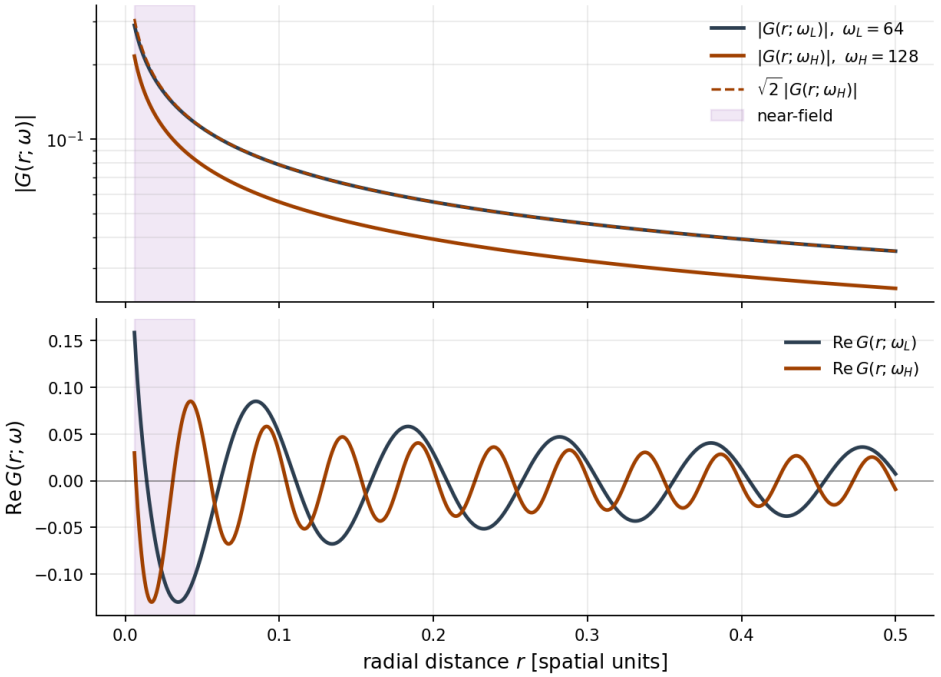


Figure 5.2: Radial cross-section of the free-space Green's function for a doubled-frequency pair. The amplitude envelope shows the far-field scaling $\sqrt{\omega_L/\omega_H}$, while the real part shows the doubled oscillation rate. A shaded near-source region marks where the far-field approximation is least accurate.

If both operators are invertible under the chosen boundary or radiation condition, the exact low-to-high solution transfer operator is

$$\mathcal{T}_{\uparrow}^{\mathrm{sol}} := \mathcal{L}_{\omega_H}^{-1}\mathcal{L}_{\omega_L}. \tag{5.7}$$

Indeed,

$$\mathcal{T}_{\uparrow}^{\mathrm{sol}} u_{\omega_L} = \mathcal{L}_{\omega_H}^{-1}\mathcal{L}_{\omega_L} u_{\omega_L} = \mathcal{L}_{\omega_H}^{-1} f = u_{\omega_H}. \tag{5.8}$$

Symmetrically, the exact high-to-low solution-transfer operator is

$$\mathcal{T}_{\downarrow}^{\mathrm{sol}} := \mathcal{L}_{\omega_L}^{-1}\mathcal{L}_{\omega_H}. \tag{5.9}$$

The superscript "sol" is important. The operator in (5.7) maps one complete solution field to another complete solution field. It satisfies the intertwining identity

$$\mathcal{L}_{\omega_H}\mathcal{T}_{\uparrow}^{\mathrm{sol}} = \mathcal{L}_{\omega_L}, \tag{5.10}$$

which expresses the fact that applying the high-frequency operator after exact transfer recovers the original source f on the solution pair.

5.2.1. The discrete defect and the residual

Chapter 2 introduced the intertwining defect $\mathcal{E}(\widetilde{\mathcal{T}})$ of an approximate transfer map $\widetilde{T}$ as the operator-level measure of how far the map departs from the ideal relation between the two frequency operators. That object is now made precise on the finite-difference grid and connected to the quantity the solver actually sees.

Work directly with the discrete operators A_{ω_L} and A_{ω_H} and the shared discrete source b, so that $A_{\omega_L}u_{\omega_L} = b$ and $A_{\omega_H}u_{\omega_H} = b$. The discrete exact up-transfer is $A_{\omega_H}^{-1}A_{\omega_L}$, the matrix counterpart of $\mathcal{T}_{\uparrow}^{\mathrm{sol}}$. For an approximate up-transfer map $\widetilde{T}$ used as a warm start, set $x_0 = \widetilde{T}u_{\omega_L}$ and define the *discrete up-transfer defect*

$$D_{\uparrow}(\widetilde{T}) := A_{\omega_L} - A_{\omega_H}\widetilde{T}. \tag{5.11}$$

The exact transfer makes this defect vanish identically, since $A_{\omega_H}(A_{\omega_H}^{-1}A_{\omega_L}) = A_{\omega_L}$ gives $D_{\uparrow}(A_{\omega_H}^{-1}A_{\omega_L}) = 0$. For an inexact map the defect is nonzero, and its action on the low-frequency solution is exactly the initial residual of the high-frequency system,

$$r_0 = b - A_{\omega_H}x_0 = A_{\omega_L}u_{\omega_L} - A_{\omega_H}\widetilde{T}u_{\omega_L} = D_{\uparrow}(\widetilde{T})\,u_{\omega_L}. \tag{5.12}$$

Equation (5.12) is the precise discrete-grid statement promised in Chapter 2, specialised to the warm-start use. Like the operator-level defect $\mathcal{E}(\widetilde{T})$, the discrete defect $D_{\uparrow}(\widetilde{T})$ vanishes for the exact transfer and grows with the approximation error. The difference is one of scope. A warm start uses the transfer map once, on a single low-frequency solution, so only the action $D_{\uparrow}(\widetilde{T})\,u_{\omega_L}$ matters, and that action is a single residual vector. A preconditioner, by contrast, would require the learned map to map an arbitrary residual to a consistent correction, so the full operator defect, rather than its action on one solution, becomes the relevant object. Section 5.4 returns to this distinction, and the preconditioner extension is taken up in Chapter 9.

The identity (5.12) also fixes the right way to read the rest of the chapter. The solver never sees the field error $e_0 = u_{\omega_H} - x_0$ directly. It sees the residual, which is the defect operator acting on the low-frequency solution. The next sections make this concrete by showing that the same field error produces very different residuals depending on which components of A_{ω_H} it excites.

5.3. Warm starts: changing the initial guess only

The most direct use of a learned solution-transfer map is as a warm start. One first solves the cheaper low-frequency problem, applies a learned transfer map $T^{\mathsf{sol}}_{\uparrow,\theta}$, and uses the result as the initial guess for the high-frequency solve:

$$x_0 = T^{\mathsf{sol}}_{\uparrow,\theta}u_{\omega_L}. \tag{5.13}$$

This use of learning is deliberately limited. The learned model chooses the initial guess x_0, but it does not change the high-frequency operator A_{ω_H}, the right-hand side b, the CSL preconditioner, or the Krylov recurrence. Once the initial guess has been set, FGMRES solves the same discrete system as it would from the zero initial guess:

$$A_{\omega_H}x = b. \tag{5.14}$$

A learned warm start can therefore improve the initial state of the iteration, but it is not itself a preconditioner.

If u_{ω_H} is the exact high-frequency solution, the initial error and residual are

$$e_0 = u_{\omega_H} - x_0, \qquad r_0 = b - A_{\omega_H}x_0 = A_{\omega_H}e_0. \tag{5.15}$$

A field-trained model is typically optimised to make the relative field error $\|e_0\|_2/\|u_{\omega_H}\|_2$ small. From a pure solution-transfer perspective, this is the right quantity: it measures how close the prediction is to the target high-frequency wavefield.

A subtlety enters as soon as the warm start is handed to an iterative solver. FGMRES is driven by the residual rather than by the field error directly, and equation (5.15) shows that these two quantities are linked through the indefinite operator A_{ω_H}. The field and the residual measure different aspects of the same prediction: the field error asks whether the predicted wavefield resembles the target solution, while the residual asks whether the prediction is a good initial state for the discrete equation. For homogeneous solution-to-solution transfer at moderate frequencies, the two views tend to agree closely. For larger frequency jumps and more complex discretisations they can diverge, and the chapter returns to this point in Sections 5.5–5.6.

For this thesis, both quantities are reported. The headline solution-transfer metric is the relative field error,

$$\frac{\|u_{\omega_H} - x_0\|_2}{\|u_{\omega_H}\|_2}, \tag{5.16}$$

and the headline solver-facing warm-start metric is the true relative residual

$$\frac{\|b - A_{\omega_H}x_0\|_2}{\|b\|_2}. \tag{5.17}$$

The first answers the natural question for a learned transfer operator. The second connects that answer to the iterative solver it is intended to accelerate.

5.3.1. The preconditioning side and the warm start

A learned warm start sits naturally inside a preconditioned Krylov method because the two act on different objects. The warm start supplies the initial state x_0; the preconditioner acts on the residuals and corrections generated once the iteration is running. This separation holds regardless of which side the CSL preconditioner is applied on. In the right-preconditioned form of equation (4.3), and equally in the left-preconditioned form $M_{\mathsf{CSL}}^{-1} A_{\omega_H} x = M_{\mathsf{CSL}}^{-1} b$, the learned x_0 enters only as the starting iterate and is never retroactively modified by the preconditioner. The decoupling of warm start and preconditioner is therefore a structural feature of the warm-start formulation, not a property of one particular side.

What the side does control is the norm the Krylov process minimises, and hence the residual the reported metrics read off. Right preconditioning solves $A_{\omega_H} M_{\mathsf{CSL}}^{-1} y = b$ and minimises the true residual $\|b - A_{\omega_H} x_k\|_2$ directly, so that quantity is available without extra work. Left preconditioning instead minimises the preconditioned residual $\|M_{\mathsf{CSL}}^{-1}(b - A_{\omega_H} x_k)\|_2$, which is the quantity the preconditioned system actually presents to the solver and, as Section 5.5 makes precise, the one more closely aligned with the field error the learned model is trained against. Under a fixed CSL factorisation the two sides differ only in this choice of minimised norm, not in the operator or the warm start, so selecting between them is a reporting and evaluation decision. The two conventions are compared directly in the one-dimensional Dirichlet setting (Section 7.4.2), and that comparison fixes the convention used for the reported experiments.

5.4. From solution transfer to preconditioning

A warm start uses frequency transfer once, before the Krylov iteration begins. A preconditioner uses a transfer mechanism repeatedly inside the iteration. This changes the mathematical task. At FGMRES iteration j, the input is a high-frequency residual $r_H^{(j)}$, and the preconditioner must return a correction $z_H^{(j)}$ such that

$$A_{\omega_H} z_H^{(j)} \approx r_H^{(j)}. \tag{5.18}$$

The output is not a complete high-frequency solution. It is an approximate error correction for the current residual.

This suggests a learned multi-frequency cycle structure, analogous to a multigrid V-cycle,

$$M_\theta^{-1} = \mathcal{P}_\theta A_{\omega_L}^{-1} \mathcal{R}_\theta, \tag{5.19}$$

where $\mathcal{R}_\theta$ maps a high-frequency residual to a low-frequency right-hand side or correction representation, $A_{\omega_L}^{-1}$ performs a low-frequency solve, and $\mathcal{P}_\theta$ prolongs the low-frequency correction back to the high-frequency system. The analogy with a geometric multigrid V-cycle is useful:

$$\text{restriction} \longrightarrow \text{coarse solve} \longrightarrow \text{prolongation}. \tag{5.20}$$

The difference is that the transfer maps must respect both grid geometry and frequency-dependent wave physics.

This learned multi-frequency cycle is not hypothetical. A closely related construction has been realised for the heterogeneous Helmholtz equation, where a U-Net is trained to act as a multigrid-style preconditioner that maps residuals to corrections inside a Krylov solver, both as a learned coarse-grid solve and as a learned deflation operator combined with shifted-Laplacian cycles [32], and later improved through compact implicit layers that address the limited receptive field of convolutional networks [33]. The present thesis takes a deliberately earlier step on the same path. Rather than constructing the in-iteration preconditioner directly, it isolates the precursor question: before a learned map can be trusted to map residuals to corrections at every iteration, what condition must a learned field prediction satisfy to be useful even once, as a warm start? The residual-amplification analysis of Sections 5.5–5.6 answers that question and explains why a field-only objective is structurally insufficient for the in-iteration role those works target.

Three transfer objects then become useful to distinguish:

$$T_{\uparrow}^{\mathsf{sol}} : u_{\omega_L} \mapsto u_{\omega_H}, \qquad \text{solution transfer}, \tag{5.21}$$

$$R : r_H \mapsto r_L \text{ or } e_L, \qquad \text{residual restriction}, \tag{5.22}$$

$$P : e_L \mapsto e_H, \qquad \text{correction prolongation}. \tag{5.23}$$

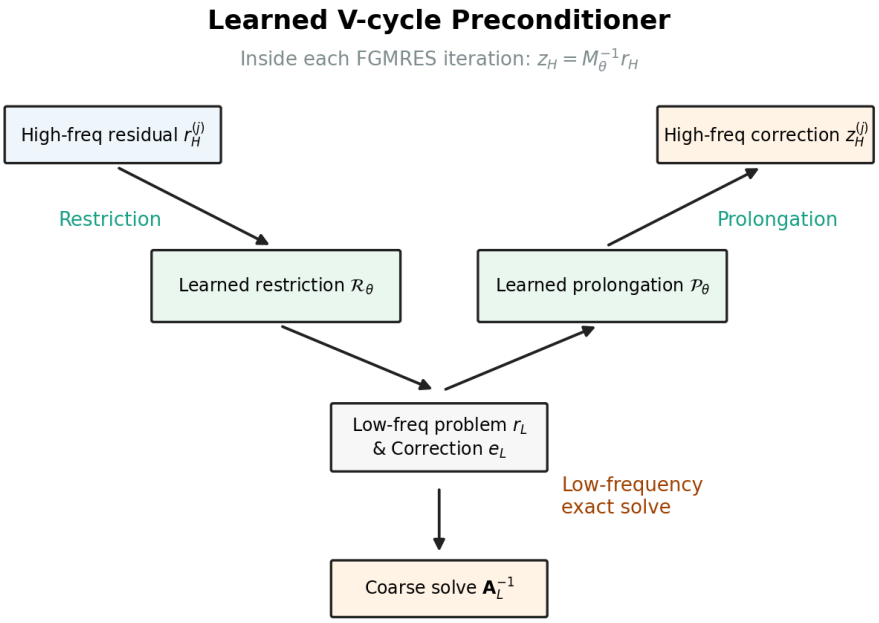


Figure 5.3: Learned V-cycle preconditioner inside one FGMRES iteration. The high-frequency residual $r_H^{(j)}$ is restricted by a learned map $\mathcal{R}_\theta$, solved with the low-frequency operator $A_{\omega_L}^{-1}$, and prolonged by a learned map $\mathcal{P}_\theta$ to produce the high-frequency correction $z_H^{(j)} = M_\theta^{-1} r_H^{(j)}$.

These objects have related physical content, but they are not interchangeable. A solution-transfer network has learned how complete solutions at two frequencies are related on the solution manifold. A residual restriction must map an arbitrary Krylov residual to a useful low-frequency problem. A correction prolongation must map an approximate low-frequency error correction to a high-frequency correction that reduces the current residual. Keeping these roles separate is what allows a single conceptual idea, frequency transfer, to support both the warm-start application of Section 5.3 and the in-Krylov preconditioning task of (5.18). The defect identity (5.12) marks the boundary between the two: the warm start is controlled by the action of the defect on a single solution, whereas the preconditioner is controlled by the defect as an operator on arbitrary residuals.

A natural question is whether these three objects are ultimately the same map under different names. They are built from the same two operators and therefore share content, but they do not collapse to one map except in a degenerate case. The cleanest way to see this is to ask what makes the learned cycle exact. Requiring $\mathcal{P} A_{\omega_L}^{-1} \mathcal{R} = A_{\omega_H}^{-1}$ pins down only the composition, not the individual factors. Two limiting choices make this explicit. If the restriction is trivial, $\mathcal{R} = I$, then exactness forces $\mathcal{P} = A_{\omega_H}^{-1} A_{\omega_L} = \mathcal{T}_\uparrow^{\text{sol}}$, so the ideal prolongation coincides with the exact up-transfer. If instead the prolongation is trivial, $\mathcal{P} = I$, then $\mathcal{R} = A_{\omega_L} A_{\omega_H}^{-1}$, which coincides with neither of the two solution-transfer operators. More generally, fixing the restriction to any invertible matrix $\mathcal{R} = S$ forces $\mathcal{P} = A_{\omega_H}^{-1} S^{-1} A_{\omega_L}$, so the factorisation admits a family of restriction–prolongation pairs, parametrised by S, that all reproduce the same target $A_{\omega_H}^{-1}$. The two limiting cases above correspond to $S = I$ and $S = A_{\omega_L} A_{\omega_H}^{-1}$. By contrast, $\mathcal{T}_\uparrow^{\text{sol}}$ and $\mathcal{T}_\downarrow^{\text{sol}}$ are each uniquely determined. The sense in which the objects are one map is thus confined to the trivial-restriction case and fails as soon as the restriction does any work.

Two further differences are decisive in the learned setting. The first is the type of object each map consumes. The solution transfer $\mathcal{T}_\uparrow^{\text{sol}}$ acts on solutions: its input $u_{\omega_L} = A_{\omega_L}^{-1} b$ lies on the low-frequency solution manifold, and a network trained on pairs $(u_{\omega_L}, u_{\omega_H})$ has only ever seen inputs of that kind. The restriction $\mathcal{R}$ acts on an arbitrary Krylov residual $r_H^{(j)}$, which is not the residual of any physical source and does not lie on that manifold, and the prolongation $\mathcal{P}$ acts on an approximate error correction rather than on a solution. A field-trained transfer map reused verbatim as a prolongation is therefore evaluated far outside

its training distribution. The second difference is the number of applications. The warm start applies the map once, so only the action of the defect on a single solution matters, as in Equation (5.12). The preconditioner applies $\mathcal{R}$ and $\mathcal{P}$ at every iteration on a sequence of shrinking residuals, so the full operator defect, and its behaviour on inputs of varying scale, becomes the relevant object. The shared algebraic origin is real, but the warm-start map and the preconditioning factors are distinct learning problems with distinct training distributions and distinct correctness conditions.

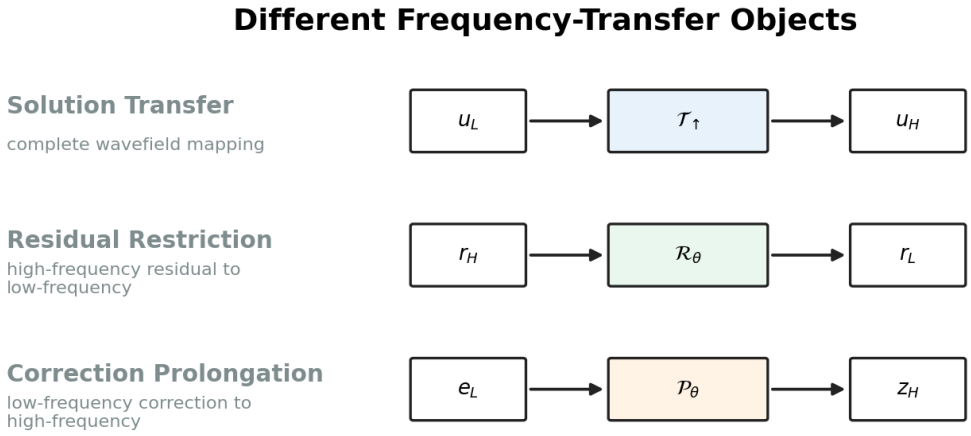


Figure 5.4: Three transfer objects that share frequency-transfer intuition but play distinct algebraic roles. Solution transfer maps one complete wavefield to another. Residual restriction maps a current high-frequency residual to a low-frequency representation. Correction prolongation maps a low-frequency correction back to the high-frequency system.

5.5. Residual amplification

The mechanism that links field error and residual error is most transparent in the one-dimensional Dirichlet problem, where the discrete Helmholtz operator admits an orthonormal eigenbasis. It is stated as a proposition so that the central identity has a precise referent in the rest of the thesis.

Proposition 5.1 (Residual amplification in the Dirichlet eigenbasis)**.** *Let $A_{\omega_H} \in \mathbb{R}^{N\times N}$ be the standard second-order finite-difference discretisation of $-\partial_{xx} - \omega^2 I$ on a uniform grid with homogeneous Dirichlet boundary conditions. Then A_{ω_H} is real symmetric and admits an orthonormal eigenbasis $\{v_k\}_{k=1}^N$ with real eigenvalues $\{\lambda_k\}_{k=1}^N$. For any warm start x_0 of the system $A_{\omega_H} u_{\omega_H} = f$, let $e_0 := u_{\omega_H} - x_0$ denote the field error and $r_0 := f - A_{\omega_H} x_0 = A_{\omega_H} e_0$ the initial true residual. Expanding $e_0 = \sum_k c_k(e_0)\, v_k$, the modal coefficients of the residual satisfy*

$$c_k(r_0) \;=\; \lambda_k\, c_k(e_0), \qquad \|r_0\|_2^2 \;=\; \sum_{k=1}^{N} |\lambda_k|^2\, |c_k(e_0)|^2. \tag{5.24}$$

Proof. A_{ω_H} is real symmetric, so by the spectral theorem it admits an orthonormal eigenbasis $\{v_k\}$ with real eigenvalues $\{\lambda_k\}$. Applying A_{ω_H} to the expansion of e_0 gives $A_{\omega_H} e_0 = \sum_k c_k(e_0)\, A_{\omega_H} v_k = \sum_k \lambda_k c_k(e_0)\, v_k$, and reading off coefficients yields $c_k(r_0) = \lambda_k c_k(e_0)$. Because $\{v_k\}$ is orthonormal, Parseval's identity, the statement that the squared Euclidean norm of a vector equals the sum of the squared magnitudes of its coefficients in an orthonormal basis, gives $\|r_0\|_2^2 = \sum_k |c_k(r_0)|^2 = \sum_k |\lambda_k|^2 |c_k(e_0)|^2$, which is the stated norm. □

The identity in Proposition 5.1 is the modal form of the defect-to-residual relation (5.12): it shows exactly how the operator redistributes a given field error across the spectrum on its way to the residual. A field-error component in a mode with small $|\lambda_k|$ contributes little to the residual, while a component of the same size in a mode with large $|\lambda_k|$ contributes a great deal. The amplification factor relating the

residual norm to the field-error norm therefore lies between $|\lambda_{\min}|$ and $|\lambda_{\max}|$, and a field-only objective does not control where in this range a given prediction falls. This is the single mechanism behind every field-versus-residual mismatch reported later; it is used as an explanatory principle rather than developed into a separate formal apparatus.

For the standard second-order Dirichlet discretisation of $-\partial_{xx} - \omega^2 I$, the eigenvectors are sine modes,

$$v_k(j) = \sqrt{\frac{2}{N+1}} \sin\left(\frac{jk\pi}{N+1}\right), \qquad j,k = 1,\dots,N, \tag{5.25}$$

with eigenvalues

$$\lambda_k(\omega) = \frac{4}{h^2} \sin^2\left(\frac{k\pi}{2(N+1)}\right) - \omega^2. \tag{5.26}$$

The eigenvalue $\lambda_k(\omega)$ in Equation (5.26) has two competing terms: a positive, monotonically increasing stiffness term from the discrete Laplacian, and a fixed negative shift $-\omega^2$ inherited from the wave equation. For small k the stiffness term is smaller than the shift, so $\lambda_k < 0$; for large k it dominates and $\lambda_k > 0$. The two terms balance near

$$k^\star \approx \frac{2(N+1)}{\pi} \arcsin\left(\frac{\omega h}{2}\right), \tag{5.27}$$

which for $\omega_H = 32$ and $N = 512$ evaluates to $k^\star \approx 10$. The spectrum therefore has two qualitatively different regions, and the residual-amplification identity $c_k(r_0) = \lambda_k c_k(e_0)$ acts differently in each. Near $k^\star$ the eigenvalues are close to zero, so these modes are nearly annihilated by the operator: they are associated with near-null behaviour and Krylov sensitivity, but they do not amplify field error into residual. Away from $k^\star$, where $|\lambda_k|$ is large, any remaining field-error component is multiplied strongly when the operator maps it into the residual. The relevant issue is therefore not a generic "resonance" of the wavefield, but the interaction between the learned error spectrum and the near-zero versus large-$|\lambda_k|$ structure of the discrete Helmholtz operator.

Figure 5.5 separates these two effects. The pink-shaded band marks the near-zero part of the spectrum, where the discrete spatial frequency nearly matches the wave number ω. Moving to the right in the figure, $|\lambda_k|$ increases, so any remaining field-error component is multiplied more strongly in the residual. The right panel shows the net effect for the learned model: although the learned field error is much smaller than the zero-start field error, its residual spectrum does not decay in the same way. This indicates that the remaining learned error is placed in operator-sensitive components.

The practical conclusion is that small field error is not enough. A learned warm start must also place its remaining error in directions that are safe for the high-frequency operator. The zero start is poor as a field approximation, but its residual spectrum is determined directly by the right-hand side. The learned start is much better as a field approximation, but its residual can be larger if the remaining error is concentrated in components that A_{ω_H} amplifies.

The amplification seen in the right panel of Figure 5.5 is therefore a property of the discrete Helmholtz operator at the chosen frequency and grid resolution, not a property of any particular learned model.

The message of Proposition 5.1 is qualitative rather than prescriptive. A learned solution transfer can be very accurate in the field while still placing a non-negligible fraction of its error in modes with large $|\lambda_k|$, where the operator converts it into a larger residual contribution. This explains why field-loss and residual metrics may disagree in detail, even when they agree on the broad ranking of models. Exploiting this mode-by-mode picture as a selection mechanism, however, is only realistic in the one-dimensional Dirichlet setting where the eigenbasis is explicit. The PML discretisations used in the rest of the thesis are complex-valued and non-normal, and they do not admit the same clean eigenvalue and eigenvector decomposition, as the next section makes precise. The proposition then fails as an equality and survives only as an inequality governed by the pseudospectrum of A_{ω_H}; Section 5.6 discusses what remains.

5.6. PML and the limits of the modal picture

The modal analysis of the previous section is exact for a Dirichlet model problem, which leads to a symmetric matrix. The solver-facing two-dimensional system in this thesis is different: it uses a finite-difference/PML operator on a full computational grid. The PML region is not a visual padding region around the physical domain; it is part of the discrete operator that the solver sees.

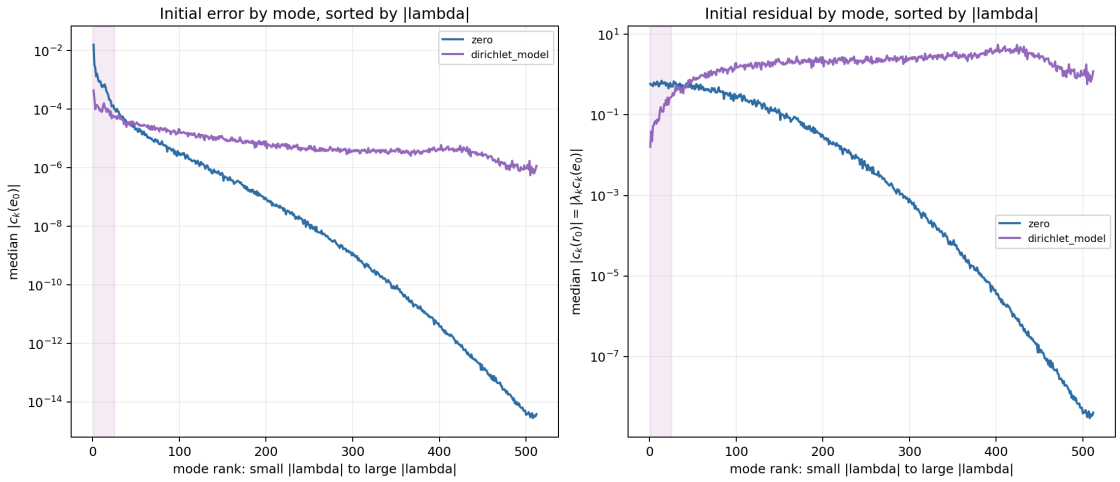


Figure 5.5: Field-error ($|c_k(e_0)|$, left) and residual ($|c_k(r_0)|$, right) coefficients in the Dirichlet eigenbasis sorted by $|\lambda_k|$ (ω_H=32, n=512; pink band: near-zero spectral band, $k^\star \approx 10$). The learned model has smaller field error but larger residual than the cold start: large-$|\lambda_k|$ modes amplify field error into residual.

Two consequences follow. First, the PML stretches coefficients into the complex plane to absorb outgoing waves, so the full-grid operator is no longer symmetric and is generally non-normal. A clean eigenvalue decomposition with real eigenvalues and orthogonal eigenvectors is not available, and the residual amplification picture cannot be read off mode by mode in the same way. The intuition that some components of the error are amplified more than others survives, but it is no longer expressed as a simple multiplication by a real eigenvalue, and there is no analogue of the explicit crossover index (5.27). The amplification is then bounded above by the operator norm $\|A_{\omega_H}\|_2$, and the directions in which it is realised are governed by the ε-pseudospectrum of A_{ω_H} rather than by its eigenvalues alone, as already noted for the CSL-preconditioned operator in Chapter 4.

Second, suppose a network predicts a high-frequency field that is accurate in the physical interior but inconsistent in the absorbing layer. The interior field may look acceptable, and the interior relative error may be small, but the full-grid residual

$$r_0 = b - A_{\omega_H} x_0 \tag{5.28}$$

is computed with the same FD/PML operator used by the solver. If the PML values are incompatible with the stencil and coefficient stretching, the residual can become large even when the physical-domain plot looks reasonable. It is therefore not generally safe to zero the PML output of a learned field, nor to train only on the physical interior and assume that the solver will ignore the absorbing layer. A solver-facing transfer model must either predict the full-grid PML field in an operator-compatible way or explicitly construct a correction whose full-grid residual is controlled.

These two observations connect the one-dimensional and two-dimensional parts of the thesis. The Dirichlet problem makes residual amplification exactly visible through eigenmodes and explains why a field-accurate prediction can still place error in a residual-sensitive band. The PML problem replaces that clean spectral picture by a complex-valued, non-normal operator on a domain that includes the absorbing layer. The diagnostic principle is the same on both sides: apply the same operator that the solver applies, and measure the residual it produces. The mode-by-mode interpretation, however, is significantly easier to compute on the one-dimensional side.

The proposition is not only a theoretical warning. It also gives an empirical test: if the learned warm start fails because its error is concentrated in operator-amplified components, then selectively removing or retaining modal components should change the initial residual in the direction predicted by (5.24). Section 7.4.1 uses this idea in the residual-gating experiment of Chapter 7. There the learned field is modified in the Dirichlet eigenbasis before being passed to FGMRES, confirming that solver usefulness depends not only on the size of the field error but on where that error lies relative to the spectrum of A_{ω_H}.

5.7. Learning objectives

The learned map studied in this thesis is first a solution-transfer map. Its training objective is therefore formulated as a field-error objective: the network is asked to reduce the discrepancy between the predicted high-frequency field and the reference high-frequency field. The residual is not the quantity minimised by the warm-start model itself. Instead, the residual is the solver-facing quantity obtained after the predicted field is inserted into the high-frequency discrete operator A_{ω_H}.

The most direct objective for solution transfer is the complex relative ℓ^2 field loss

$$\mathcal{L}_{\text{field}} = \frac{\|\widehat{u}_{\omega_H} - u_{\omega_H}\|_2^2}{\|u_{\omega_H}\|_2^2 + \varepsilon}. \tag{5.29}$$

Here $\varepsilon > 0$ is a small stabilisation constant. The real and imaginary parts of $\widehat{u}_{\omega_H}$ and u_{ω_H} are represented as separate channels in the network, but the norm is interpreted as the norm of the corresponding complex vector. This is the natural objective for learning $T_{\uparrow}^{\text{sol}}$: it answers whether the low-frequency field contains enough information to predict the high-frequency wavefield. It is the only objective used to train the models reported in this thesis.

The residual-amplification mechanism of Section 5.5 suggests two further objectives that align the loss with the quantity the solver sees. They are defined here because they delimit the design space for the residual-aware extension discussed in Chapter 9; they are not trained in the present experiments, and the chapter is explicit about this so that the field-loss results are not misread as residual-trained results.

A residual-based objective becomes natural when the learned component is used as a correction model inside the solver. In that setting the network is not asked to predict the complete high-frequency solution. It is asked to predict an error correction $\widehat{e}$ for a given residual r, so that

$$A_{\omega_H}\widehat{e} \approx r. \tag{5.30}$$

The corresponding operator-aligned loss is

$$\mathcal{L}_{\text{op}} = \frac{\|A_{\omega_H}\widehat{e} - r\|_2^2}{\|r\|_2^2 + \varepsilon}. \tag{5.31}$$

This loss should be interpreted as an error objective in the equation induced by the residual, not as a replacement for Krylov residual minimisation. GMRES or FGMRES still performs the residual minimisation over the Krylov subspace. The role of $\mathcal{L}_{\text{op}}$ is to train the learned component to produce corrections that are compatible with the operator applied by the solver.

A third option measures residual error in a preconditioned norm. If M_{CSL}^{-1} is a trusted baseline preconditioner, such as the complex shifted Laplacian, then

$$\mathcal{L}_{\text{prec}} = \frac{\|M_{\text{CSL}}^{-1}(A_{\omega_H}\widehat{e} - r)\|_2^2}{\|M_{\text{CSL}}^{-1}r\|_2^2 + \varepsilon} \tag{5.32}$$

asks whether the learned correction is useful in the norm seen by the preconditioned Krylov process. This is not identical to the true residual, but it can be a meaningful training or diagnostic quantity when the learned method is intended to augment a baseline preconditioner.

The later chapters use these distinctions to organise the experimental evidence. The primary scientific question is whether learned solution transfer is accurate enough to be a useful initial state for the high-frequency solve; the residual-based objectives are the lens through which the same field-trained predictions are evaluated as solver components, and they set the agenda for the residual-aware training direction rather than being optimised here.

5.8. Implications for the learned model

The analysis above determines what the learned model must be evaluated on, but not a unique architecture. It implies three requirements: the model must preserve complex phase information, use enough spatial context to transform frequency-dependent oscillations, and be judged by the residual produced by

Table 5.1: Learning objectives and the mathematical question each one answers. Only the field loss is used to train the models in this thesis; the two residual-aligned objectives define the design space for the residual-aware preconditioning direction of Chapter 9.

Objective	Question answered	Status in this thesis
Field loss	Does the predicted wavefield match the target solution?	Used to train all reported models
Operator-aligned loss	Does the predicted correction solve $A_{\omega_H} e = r$?	Defined; deferred to future work
Preconditioned residual loss	Does the correction help in the baseline preconditioned norm?	Defined; deferred to future work

A_{ω_H}, not only by visual or field-error quality. The concrete architecture, input channels, training loss, and solver-evaluation protocol are therefore specified in Chapter 6.

A learned solution map of this kind is one instance of the broader programme of learning maps between function spaces, in which neural operators are trained to approximate parametric solution operators of partial differential equations [14]. The transfer map studied here is a restricted special case: it is tied to a fixed grid, a fixed medium, and a single factor-of-two frequency pair, and it is evaluated not as a standalone surrogate but by the residual it induces in the high-frequency solver. This restriction is deliberate. It keeps the object simple enough to diagnose mode by mode in the Dirichlet setting, which is what makes the residual-amplification mechanism visible.

This thesis focuses on U-Net transfer models because they are expressive enough to learn frequency-coherent solution maps, while remaining simple enough to diagnose. The scientific question is not whether a larger neural operator can reduce field loss further. The question is whether learned frequency transfer is sufficiently accurate to act as a useful warm start, and what remains to be done before the same building block can be turned into a residual-compatible solver component.

5.9. Summary

This chapter introduced the frequency transfer operator as the mathematical link between low- and high-frequency Helmholtz fields. The Green's function shows why such transfer is physically plausible: changing frequency modifies phase and amplitude scale while preserving propagation geometry. Exact solution transfer is therefore well-defined as $\mathcal{T}_{\uparrow}^{\mathrm{sol}} = \mathcal{L}_{\omega_H}^{-1}\mathcal{L}_{\omega_L}$ and maps the low-frequency solution manifold onto the high-frequency one. The discrete intertwining defect $D_{\uparrow}(\widetilde{T}) = A_{\omega_L} - A_{\omega_H}\widetilde{T}$ measures the failure of an approximate map, and its action on the low-frequency solution is exactly the initial residual seen by the solver.

The chapter then traced the two routes by which such a learned operator can be used. As a warm start, it provides an initial guess for the high-frequency solve and is naturally judged by the relative field error. As a building block of a learned V-cycle, it must be decomposed into restriction and prolongation maps that act on residuals and corrections rather than on complete solutions. The two roles share frequency-transfer intuition but lead to different mathematical tasks, separated by whether the defect acts on a single solution or on arbitrary residuals.

A final refinement linked the two views. Because $r_0 = A_{\omega_H}(u_{\omega_H} - x_0)$, small field error can correspond to larger residual error in operator-amplified components. The one-dimensional Dirichlet model makes this exact and predicts the location of the dominant amplification band from the crossover index. The PML setting used elsewhere in the thesis preserves the intuition but not the clean eigenstructure. This sets up the experimental chapters that follow, where learned solution transfer is evaluated both as a wavefield prediction and as an initial state for the high-frequency Helmholtz solve.

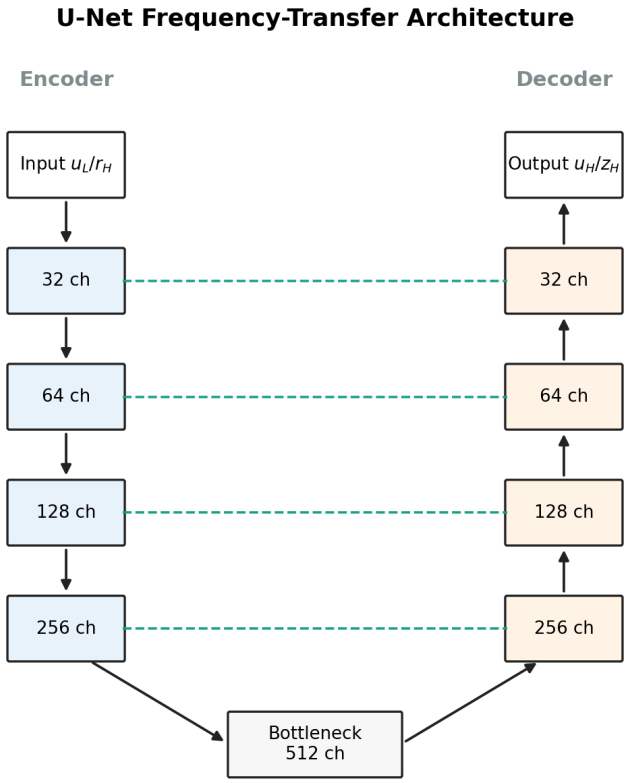


Figure 5.6: U-Net as a frequency-transfer model. The encoder aggregates large-scale wavefield structure, while skip connections preserve oscillatory detail needed by the decoder. The same architectural family can be used for solution transfer or correction prediction, but the target and loss determine which mathematical object is learned.

6

Experimental design and computational setup

Chapter 5 defined the solution-transfer operator and showed why a learned high-frequency field must be judged by the residual induced by the discrete Helmholtz operator, not only by relative field error. If

$$e_0 = u_{\omega_H} - \hat{u}_{\omega_H}, \qquad r_0 = b - A_{\omega_H}\hat{u}_{\omega_H} = A_{\omega_H}e_0,$$

then the solver sees the operator-applied error $A_{\omega_H}e_0$, not the field error e_0 itself. In the Dirichlet diagnostic setting this amplification is exact mode by mode: each field-error coefficient is multiplied by the corresponding eigenvalue before it appears in the residual. A visually accurate prediction can therefore still be a poor algebraic initial state if its error lies in residual-sensitive components. This threshold is referred to as the residual-amplification barrier.

This chapter translates that theoretical obstacle into a concrete experimental programme. The predicted field is inserted as the FGMRES initial guess, and success is measured by whether the initial true residual and the subsequent FGMRES trajectory improve relative to a zero start. A positive result shows that learned transfer supplies solver-useful information. A negative result remains informative when it identifies the residual-amplification barrier, PML incompatibility, or both.

The chapter first states the design principles and experimental ladder, then specifies the discrete systems, datasets, learned models, solver protocol, metrics, verification checks, and reproducibility conventions.

6.1. Design principles

The experimental programme follows five principles.

1. Separate field prediction from solver usefulness. The neural network is trained to predict a high-frequency wavefield, but the solver is controlled by residuals. The primary experimental question is therefore not only whether $\|\hat{u}_{\omega_H} - u_{\omega_H}\|_2$ is small, but whether $\|b - A_{\omega_H}\hat{u}_{\omega_H}\|_2$ is small enough to improve the subsequent Krylov solve. All experiments report both quantities. This prevents a visually successful field predictor from being mistaken for a useful solver component.

2. Use a diagnostic ladder rather than a single benchmark. The experiments move through three settings. The one-dimensional Dirichlet problem provides an exact spectral microscope because the eigenvectors of the operator are known analytically. The one-dimensional PML problem adds the absorbing-layer compatibility issue while remaining small enough to inspect. The two-dimensional FD/PML problem is the solver-facing setting of the thesis. This ladder strengthens the interpretation: a failure in two dimensions can be traced back to modal residual amplification, PML incompatibility, or both.

3. Keep the algebraic problem fixed within each comparison. When a learned warm start is compared with a zero start, the frequency, grid, PML, right-hand side, CSL preconditioner, damping parameter, stopping criterion, and FGMRES implementation are held fixed. Only the initial guess x_0 changes. This isolates the algebraic effect of the learned transfer model. The computational cost of obtaining x_0 is accounted for separately in Section 6.6.2, because a learned warm start requires a low-frequency solve and model evaluation while the zero start does not.

4. Evaluate on the same operator family used by the solver. Earlier field-only workflows can give misleadingly good interior predictions while producing PML values that are incompatible with the full FD/PML operator. The final two-dimensional experiments therefore use exact FD/PML data, train on full-grid targets, and evaluate residuals with the same operator used by FGMRES. The PML is treated as part of the algebraic system, not as a visual border.

5. Treat negative results as tests of theory. The residual-amplification barrier predicts that a learned warm start must achieve extremely small residual-compatible error before it can beat the zero-start in initial true residual. If a learned model fails this threshold, the result is not dismissed as an implementation failure. It is compared with the theoretical barrier, the modal diagnostics, and the PML diagnostics. This is important for scientific significance: the experiments test a mechanism, not only a performance claim.

6.2. Experimental ladder

The experimental design uses three increasingly realistic problems. Each setting introduces a new difficulty while keeping the rest of the setup as similar as possible.

One-dimensional Dirichlet problem. The first setting is the homogeneous one-dimensional Helmholtz equation with Dirichlet boundary conditions and no absorbing layer. Its eigenpairs are analytical, so field error and residual error can be compared mode by mode. This is the only setting in which the residual-amplification prediction can be tested exactly.

One-dimensional PML problem. The second setting keeps the dimension low but replaces the Dirichlet boundary by a PML. The simple sine-basis analysis no longer applies because the PML operator is complex-valued and non-normal. The role of this setting is to isolate absorbing-layer compatibility: a learned field may be accurate in the physical interior while producing a large full-grid residual if its PML values are inconsistent with the discrete stencil.

Two-dimensional FD/PML problem. The final setting is the solver-facing problem of the thesis: a two-dimensional finite-difference Helmholtz operator on a 512×512 grid with a PML of width 112 cells per side and a 288×288 physical interior. All residuals are computed on the full grid using the same operator that FGMRES applies. Interior field errors are still reported as diagnostics, but the full-grid residual is the decisive algebraic quantity.

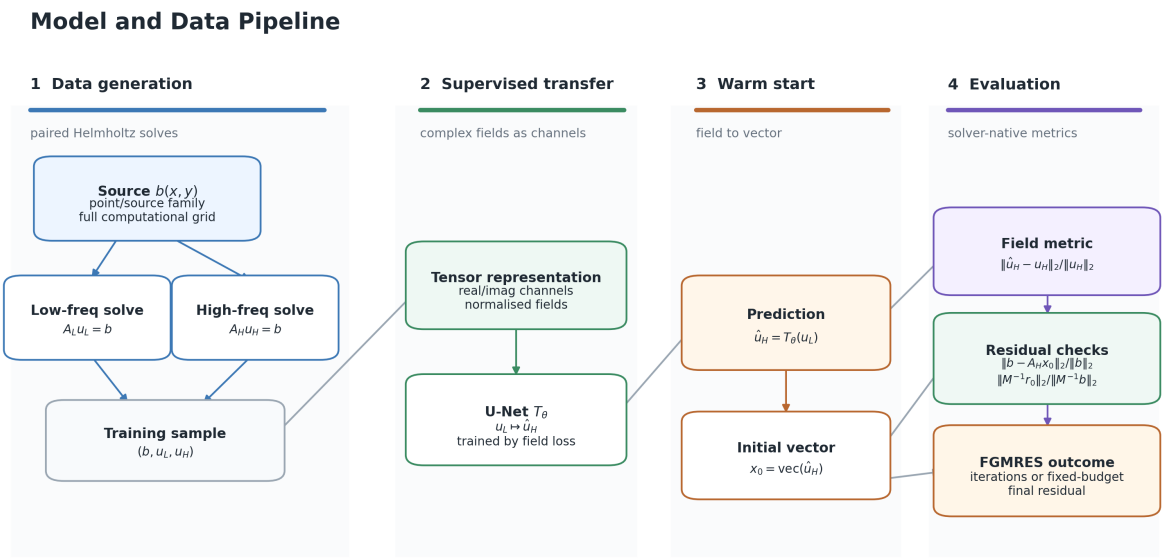


Figure 6.1: Overview of the experimental pipeline.

6.3. Discrete Helmholtz systems

All experiments solve linear systems of the form

$$A_\omega u_\omega = b, \tag{6.1}$$

with the residual convention

$$r(x) = b - A_\omega x. \tag{6.2}$$

This convention is used for training diagnostics, warm-start evaluation, and FGMRES convergence reporting throughout the thesis.

6.3.1. One-dimensional Dirichlet problem

The clean diagnostic problem uses 512 unknowns and no absorbing layer. The tested frequency pair is

$$\omega_L = 16, \qquad \omega_H = 32. \tag{6.3}$$

With the one-dimensional sign convention

$$A_\omega = -D_{xx} - \omega^2 I, \tag{6.4}$$

the eigenpairs are analytical:

$$\lambda_k(\omega) = \frac{4}{h^2}\sin^2\left(\frac{\pi k}{2(N+1)}\right) - \omega^2, \qquad v_k(j) = \sqrt{\tfrac{2}{N+1}}\sin\left(\tfrac{jk\pi}{N+1}\right), \tag{6.5}$$

for $j, k = 1, \ldots, N$. The eigenvectors satisfy $\|v_k\|_2 = 1$, so the prefactor $\sqrt{2/(N+1)}$ is the Euclidean normalisation constant and the basis $\{v_k\}$ is orthonormal.

This spectral structure makes the field-error/residual-error relation explicit. If $u_\star$ is the high-frequency solution, x_0 is an initial guess, and $e_0 = u_\star - x_0$, then

$$r_0 = b - A_{\omega_H} x_0 = A_{\omega_H} e_0, \qquad c_k(r_0) = \lambda_k\, c_k(e_0). \tag{6.6}$$

This identity turns the field-error/residual-error distinction from a qualitative warning into an exact mode-by-mode diagnostic, which is the purpose of the Dirichlet case.

6.3.2. One-dimensional PML problem

The one-dimensional PML problem uses the same frequency pair, $16 \to 32$. The grid is chosen as a one-dimensional slice of the 2D FD/PML grid: a physical interior of 288 cells and a PML of width 112 cells on each side, giving 512 unknowns. This choice size-matches the 1D Dirichlet problem and dimension-matches the 2D FD/PML interior.

The symmetric sine-basis analysis of Section 6.3.1 no longer applies because the PML operator is complex-valued and generally non-normal. The role of this experiment is therefore narrower: it tests whether a learned field is compatible with the PML stencil. A predicted field can be accurate in the physical region while producing a large full-grid residual if the PML values are inconsistent with the discrete stencil and stretching coefficients. The 1D PML experiment therefore reports full-grid field errors, true residuals, preconditioned residuals, and FGMRES curves, but does not present the PML problem as an exact modal decomposition.

6.3.3. Two-dimensional FD/PML problem

The solver-facing experiments use a two-dimensional finite-difference Helmholtz operator on a 512×512 computational grid. A PML of width 112 grid points surrounds the physical domain on each side, leaving a 288×288 interior. The tested frequency pairs are

$$16 \to 32, \qquad 32 \to 64, \qquad 64 \to 128. \tag{6.7}$$

The PML is part of the algebraic system, not a removable visual border. All solver-facing residuals are computed on the full 512×512 grid using the same FD/PML operator used by FGMRES. Interior metrics are reported as additional diagnostics. The primary question is whether the predicted full-grid field gives a useful state for the original high-frequency equation.

Table 6.1: Discrete systems used in the experimental ladder. The frequency pairs are $16 \to 32$ for the two 1D diagnostics and $16 \to 32$, $32 \to 64$, $64 \to 128$ for the 2D FD/PML solver-facing experiments.

Setting	Grid	Boundary treatment	Role
1D Dirichlet	512	Dirichlet, no PML	exact modal diagnostic
1D PML	512	PML 112 per side	absorbing-layer diagnostic
2D FD/PML	512×512	PML 112 per side	solver-facing evaluation

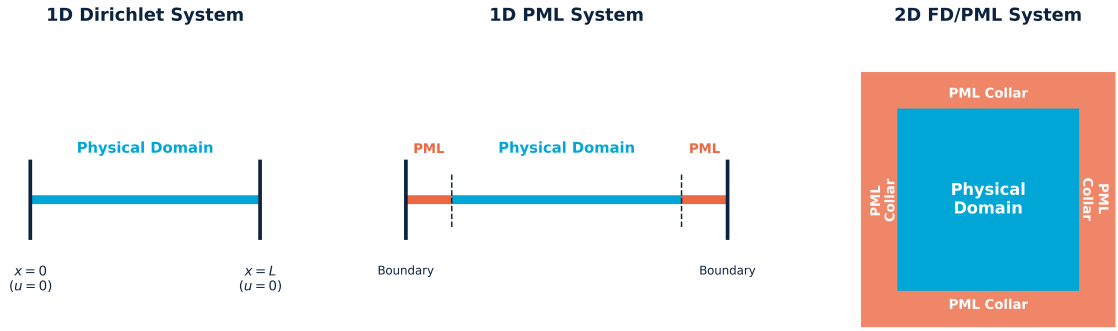


Figure 6.2: The three settings of the experimental progression. Each panel indicates the physical domain, the absorbing layer where present, and the operator used to compute the residual. The Dirichlet problem is real-valued and self-adjoint; the two PML problems are complex-valued and non-normal.

6.4. Data generation and storage convention

Each supervised sample is a triple $(b,\, u_{\omega_L},\, u_{\omega_H})$: a source field, the corresponding low-frequency solution, and the corresponding high-frequency solution. The same source is used for the low- and high-frequency solves, so that the supervised pair $(u_{\omega_L}, u_{\omega_H})$ is generated by a single forcing under two operators.

6.4.1. Source model for the 2D FD/PML datasets

Each right-hand side is constructed as a superposition of K two-dimensional Gaussian bumps with complex amplitudes,

$$b(x,y) = \sum_{j=1}^{K} \alpha_j \exp\left(-\frac{(x-x_j)^2 + (y-y_j)^2}{2\sigma_j^2}\right), \qquad \alpha_j = a_j\, e^{i\phi_j}, \tag{6.8}$$

with the following sampling convention. The number of bumps K is drawn uniformly at random from $\{3,4,5,6\}$. Each width is fixed at $\sigma_j = 2$ grid cells. Each magnitude is $a_j \sim \mathcal{U}(1,2)$, and each phase is $\phi_j \sim \mathcal{U}(0,2\pi)$. Each centre (x_j, y_j) is sampled uniformly on the integer physical interior grid $\{112,113,\ldots,399\}\times\{112,113,\ldots,399\}$. The complex source b is passed in full to both the low-frequency and high-frequency solves.

After the solves, every sample is divided by the per-sample normalisation factor

$$\rho \;=\; \sqrt{|u_{\omega_L}|^2_{\text{interior}}} \;+\; \varepsilon, \qquad \varepsilon = 10^{-8}, \tag{6.9}$$

the root-mean-square of $|u_{\omega_L}|$ over the physical interior, before storage. The stored arrays are therefore in dimensionless, normalised units; the factor ρ is retained in the auxiliary array `rms.npy` so that predictions can be returned to physical units when needed. The imaginary part of the source is constructed and used in the solves, but only the real part of the normalised source is written to disk in the array `source_re.npy`. The solutions u_{ω_L} and u_{ω_H} are stored as separate real and imaginary channels because both are required to evaluate the complex residual.

6.4.2. Dataset specification

The main two-dimensional dataset is generated under random seed 42 for the up-transfer direction. It contains 9600 samples per frequency pair for the three pairs in Equation (6.7), for a total of 28800 samples. The operational dataset name, array paths, and artefact locations are listed in Appendix 10.

Sample ordering on disk is block-contiguous by frequency pair: samples 0 through 9599 are $16 \to 32$, samples 9600 through 19199 are $32 \to 64$, and samples 19200 through 28799 are $64 \to 128$. This is exploited at train time by indexing into the corresponding slice when a single-pair experiment is required.

The solver-facing models reported in this thesis are trained per frequency pair, each on the block of 9600 samples belonging to its pair. Joint training across all three pairs was an early exploratory configuration and is not used for any reported solver result. The block-contiguous ordering above is what makes this per-pair selection straightforward.

Two split conventions are used, sharing the seed 42 but not the ratios. The pooled hyperparameter sweep splits the full dataset of 28800 samples into training, validation, and held-out test sets in proportions $70\,\%/15\,\%/15\,\%$, applied by a global random permutation before the split. This yields $20\,160$ training, $4\,320$ validation, and $4\,320$ test samples, and the permutation ensures that each split contains a mixture of all three frequency pairs in approximately equal proportions. The per-pair solver models reported in Chapter 7 are instead trained on their own 9600-sample block under a random train/validation split with validation fraction 0.2. In both conventions, training data determine network weights and validation data determine checkpoint selection and hyperparameter choices; the two are recorded together in Appendix 10.

Alongside the main FD/PML dataset, a second exploratory dataset was generated in which the complex source is stored explicitly and the paired solutions are computed from the analytic free-space Green's function rather than from the FD/PML solver. It exists for one specific purpose: the source-conditioning study of Section 7.3, which asks whether a network that is given the source learns a genuine cross-frequency transfer or merely a direct source-to-solution map. Because its solutions satisfy the free-space Green's relation and not the discrete FD/PML system, the residual it produces under the FD/PML operator is not comparable with the residuals of the solver-facing experiments. It is therefore excluded from every warm-start claim in Chapters 7 and 8, and its metadata records the analytic origin so that the two dataset families cannot be combined by accident.

Table 6.2: Storage convention for the main 2D FD/PML dataset. All grid arrays are stored as `float32` in normalised units. The per-sample normalisation factor is stored in `rms.npy`.

Array	Shape	Role
`source_re`	$(28800, 512, 512)$	real part of normalised complex source
`u_low_re`	$(28800, 512, 512)$	low-frequency solution, real part
`u_low_im`	$(28800, 512, 512)$	low-frequency solution, imaginary part
`u_high_re`	$(28800, 512, 512)$	high-frequency solution, real part
`u_high_im`	$(28800, 512, 512)$	high-frequency solution, imaginary part
`omega_low`	$(28800,)$	low-frequency angular frequency ω_L
`rms`	$(28800,)$	RMS normalisation factor ρ

6.4.3. Data for the 1D problems

The two 1D problems share the source model and discretisation parameters of the 2D pipeline. Each 1D dataset uses $K \in \{3, \dots, 6\}$ complex Gaussian bumps of width $\sigma = 2$ grid cells, amplitudes $a_j \sim \mathcal{U}(1, 2)$ with random phases, placed uniformly in the physical interior $[112, 400)$. Each dataset contains 2400 paired samples on a grid with 512 unknowns, generated under random seed 42. The split convention is sequential rather than randomised: the first 2000 samples are training, the remaining 400 are validation. The benignness of this non-shuffled split is verified in Section 6.8.

Four distinct 1D datasets exist, corresponding to the analytical and discrete solver families used in the thesis.

Table 6.3: One-dimensional datasets used in the thesis ("corrected flux-PML" is the flux-differenced absorbing stencil of Section 3.7 after a sign and scaling fix). The exact artefact paths are listed in Appendix 10.

Role	Solver convention	Pair(s)	Used for
Green reference	free-space Green function	all three	saturation baseline
FD/PML 1D	finite-difference PML	all three	PML diagnostics
Corrected flux-PML	corrected PML flux form	$16 \to 32$	main 1D PML comparison
Dirichlet diagnostic	closed Dirichlet operator	$16 \to 32$	modal residual analysis

The Green's function datasets supply a saturating baseline: they correspond to the analytically solvable free-space problem and provide a reference against which discrete-solver effects can be isolated. The FD/PML 1D datasets supply paired data for the absorbing-layer setting. The corrected flux-PML dataset is the variant used in the main 1D PML comparison. The term "flux-PML" refers to the conservative, flux-differenced form of the absorbing stencil already written in Equation (3.12), in which the complex stretching factor $s(x) = 1 + i\sigma(x)/\omega$ enters through the face-averaged fluxes $s_{i\pm1/2}$ rather than through a single nodal coefficient. "Corrected" records a concrete fix relative to the first FD/PML implementation: an earlier version applied the stretching with an inconsistent sign and scaling on those face terms, which left the assembled operator slightly inconsistent with the residual convention $r(x) = b - A_\omega x$; the corrected variant restores that consistency. The name is introduced here only to keep the four 1D datasets distinguishable, and the underlying stencil is the one already derived in Section 3.7. The canonical 1D Dirichlet dataset is the one used for the modal-identity diagnostics of Section 6.3.1.

6.5. Learned models and training objectives

The baseline learned object is a solution-transfer map

$$T^{\mathsf{sol}}_{\uparrow,\theta} : u_{\omega_L} \mapsto \hat{u}_{\omega_H}, \tag{6.10}$$

trained with a complex relative ℓ^2 field loss

$$\mathcal{L}_{\mathsf{field}} = \frac{\|\hat{u}_{\omega_H} - u_{\omega_H}\|_2^2}{\|u_{\omega_H}\|_2^2 + \varepsilon}. \tag{6.11}$$

Earlier exploratory models evaluated this loss primarily on the physical interior; the final models evaluate it on the full grid so that the learned output includes the absorbing layer and is therefore directly insertable into the FD/PML system.

6.5.1. Input tensor representation

The pipeline of Figure 6.1 represents every field as a real tensor before it enters a network, but the exact channel layout is not fixed across the experiments and is stated per model in Table 6.4. Three conventions are shared by all models. A complex field is always carried as two real channels, its real and imaginary parts, rather than as an amplitude and phase pair, for the reason given in Section 6.5.2. Every field is presented in the per-sample normalised units of Equation (6.9). The spatial shape is the full computational grid, 512 in one dimension and 512×512 in two, so the absorbing layer is always part of the tensor and is never cropped before the network sees it.

What varies is the number and meaning of the conditioning channels appended to the two field channels. The two-dimensional `TransferUNet` adds only static geometric and frequency conditioning, namely a PML ramp, a meshgrid coordinate pair, and a broadcast ω_L channel, and it does not receive the source. The one-dimensional `TransferUNet1d` additionally receives the real and imaginary parts of the source, a single coordinate channel, and the broadcast ω_L. The dilated CNN receives only the two field channels. The analytic diagnostic model of Section 7.3 uses a wider 29-channel encoding that includes Fourier coordinate features. The differences are deliberate: source channels are withheld from the solver-facing two-dimensional model so that it learns frequency transfer rather than direct source-to-solution synthesis, whereas the one-dimensional diagnostic models are given extra inputs precisely to test whether that information changes residual behaviour once the field prediction is already accurate.

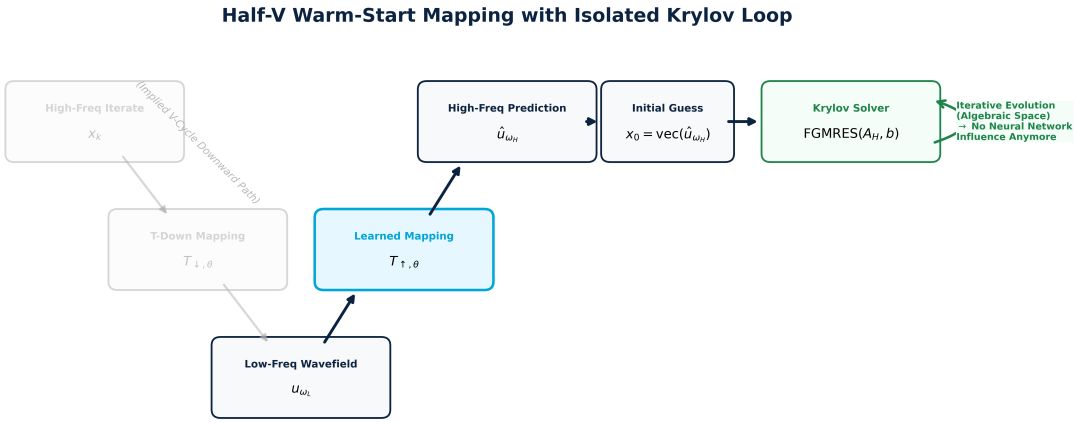


Figure 6.3: The learned warm-start mapping. A low-frequency field u_{ω_L} is mapped by the network $T_{\uparrow,\theta}$ to a predicted high-frequency field $\hat{u}_{\omega_H}$, which is then inserted as the FGMRES initial guess x_0. The same operator A_{ω_H} that produced u_{ω_H} during data generation is used to compute the initial residual $r_0 = b - A_{\omega_H} x_0$.

6.5.2. Architectural requirements

The frequency-transfer analysis of Chapter 5 imposes four requirements on the learned model.

First, the model needs a nontrivial receptive field. The phase transformation from ω_L to ω_H depends on propagation distance and cannot be represented by independent pointwise scaling. Convolutional architectures are therefore natural because they aggregate neighbouring wavefield information through stacked local operations.

Second, the representation must preserve complex phase. This thesis represents complex fields by separate real and imaginary channels. An amplitude–phase representation is possible, but it introduces discontinuities at phase wraps, which are inconvenient for standard convolutional networks.

Third, the model needs multiscale structure. Low-frequency fields carry broad propagation envelopes, while high-frequency targets require oscillatory detail. This motivates a U-Net architecture: the encoder aggregates large-scale context, the decoder reconstructs fine-scale structure, and skip connections preserve spatial detail.

Fourth, architecture alone does not guarantee solver usefulness. A network can have enough expressive power to predict plausible fields and still place its remaining error in residual-sensitive components. The solver-facing evaluation therefore reports true residuals and FGMRES curves alongside the field-error metrics used during training.

6.5.3. One-dimensional models: a comparative pair

Two architectures are trained on the canonical 1D Dirichlet dataset: a dilated CNN and a one-dimensional U-Net. Both predict u_{32} from u_{16}, both are trained on the same paired data with the same optimiser, and both use the same complex relative ℓ^2 field loss. The comparison is not a perfectly controlled architecture ablation, because the U-Net also receives source channels and static conditioning. It is instead a matched diagnostic pair: once both models reach sub-percent field error, their residual behaviour tests whether architecture and input information materially affect solver usefulness.

Dilated CNN. The dilated CNN is a residual stack of 8 dilated 1D convolutions of width 128 with ReLU activations, selected through an Optuna hyperparameter search. Its best validation field loss is 5.37×10^{-3}, reached at epoch 488 of a 500-epoch budget.

One-dimensional U-Net (`TransferUNet1d`). The 1D U-Net is a direct one-dimensional port of the 2D `TransferUNet`. It is a five-level encoder–decoder with `base_ch` $= 32$, giving the channel-width progression $32 \to 64 \to 128 \to 256 \to 512$. Architectural primitives match the 2D variant with `Conv2d` replaced by `Conv1d` and bilinear upsampling replaced by linear upsampling.

The 1D U-Net receives the real and imaginary parts of u_{ω_L}, the real and imaginary parts of the source b, a PML ramp, a normalised spatial coordinate, and a broadcast ω_L channel. This model is a diagnostic instrument rather than a production warm-start model: unlike the solver-facing 2D `TransferUNet`, which is deliberately denied the source, the 1D U-Net is given b precisely to test whether supplying source information changes the residual behaviour once the field prediction is already accurate. The source channels are therefore an experimental control, not a component of any warm start whose residual is reported as a solver result. The best validation field loss of this network is 4.34×10^{-3}, reached at epoch 221.

This model pair is used in three diagnostic roles. First, the raw prediction $\hat{u}_{32}$ is inserted as the FGMRES initial guess and the initial true residual is recorded. Second, the prediction is decomposed in the known Dirichlet eigenbasis, and individual modes are filtered or gated. Third, the field-trained map is used inside preconditioning-style experiments that separate residual restriction, low-frequency solving, and high-frequency correction prolongation. These are diagnostic tests of whether field-trained information can be made residual-safe.

The 1D PML setting uses learned warm-start variants from the corrected flux-PML dataset. The principal architectural choice in that comparison is whether the PML output is kept, zeroed at evaluation time, or trained as part of the full operator-compatible field.

6.5.4. Two-dimensional model

The main two-dimensional architecture is a five-level `TransferUNet` with `base_ch` $= 32$, giving the same channel-width progression $32 \to 64 \to 128 \to 256 \to 512$ as the 1D U-Net. The input is the real and imaginary parts of the low-frequency field, augmented by static 2D PML ramp and meshgrid conditioning channels and a per-sample ω_L broadcast. The target is the real and imaginary parts of the high-frequency field on the full 512×512 FD/PML grid. The loss is the complex relative ℓ^2 field loss in Equation (6.11), evaluated on the full grid.

The source field b is withheld from the 2D model, and this is a substantive decision rather than a dimensional accident. The 2D `TransferUNet` is the only model whose prediction is inserted into FGMRES as a warm start, so it alone carries the solver claims of the thesis. For that model the question under test is strictly whether the high-frequency field can be recovered from the low-frequency one, $u_{\omega_L} \mapsto u_{\omega_H}$. Because the high-frequency solution also satisfies $A_{\omega_H} u_{\omega_H} = b$, a model that is given b could in principle bypass the transfer task and learn the source-to-solution map $b \mapsto u_{\omega_H}$ directly. That is a different and easier problem, and a warm start produced that way could no longer be attributed to cross-frequency information. Withholding b removes that escape route, so the reported solver behaviour is attributable to frequency transfer alone. The 1D `TransferUNet1d` is given b for the opposite, deliberately diagnostic reason: it carries no solver claim, and feeding it the source is the control that measures whether source information changes residual behaviour once the field prediction is already accurate. The source-free dilated CNN supplies the matching source-free 1D point. The three models therefore bracket the source question, with the production model placed deliberately on the source-free side.

The final experimental comparison distinguishes four learned-start variants:

cold	$x_0 = 0$,
interior model, PML zeroed	$x_0 = \hat{u}_{\omega_H}$ with the PML region set to zero,
flux-full, raw	$x_0 = \hat{u}_{\omega_H}$ on the full FD/PML grid,
flux-full, PML zeroed	$x_0 = \hat{u}_{\omega_H}$ with the learned PML region removed.

(6.12)

This comparison tests whether the PML region produced by the full-grid model is useful to the solver, or whether it should be discarded at evaluation time.

6.6. CSL–FGMRES solver protocol

The complex shifted Laplace preconditioner M_{CSL} is part of the solver, not part of the operator and not part of the metric. The operator A_{ω_H} is the discrete high-frequency Helmholtz matrix; it does not change. The Krylov method is FGMRES [13]. Flexible preconditioning is defined on the right, and the right-preconditioned form is the natural host for the learned extensions of Chapter 9; with the fixed

Table 6.4: Learned models used in the experiments. All are trained as field predictors with the same complex relative ℓ^2 loss; solver usefulness is tested separately by residuals and FGMRES curves.

Setting	Model	Input information	Target
1D Dirichlet	Dilated CNN	u_{ω_L} (Re, Im)	full-grid u_{ω_H}
1D Dirichlet	`TransferUNet1d`	u_{ω_L}, source b, PML ramp, coordinate, ω_L	full-grid u_{ω_H}
1D PML	corrected flux-PML variants	u_{ω_L}-based channels	full/PML-dependent u_{ω_H}
2D FD/PML	`TransferUNet`	u_{ω_L}, PML ramp, meshgrid, ω_L	full-grid u_{ω_H}

CSL factorisation used here, the left-preconditioned iteration reduces to standard GMRES applied to $M_{\mathrm{CSL}}^{-1}A_{\omega_H}x = M_{\mathrm{CSL}}^{-1}b$. Both insertions are implemented, and the one-dimensional Dirichlet study compares them explicitly (Section 6.6.1). On the basis of that comparison, the reported experiments from Section 7.4.2 onward use left preconditioning, because the preconditioned residual is the quantity most faithful to the field objective the learned model is trained against. Three metrics are reported throughout, in the hierarchy fixed in Chapter 7: the field error as the scientific target, the preconditioned residual $\|M_{\mathrm{CSL}}^{-1}(b - A_{\omega_H}x_k)\|_2/\|M_{\mathrm{CSL}}^{-1}b\|_2$ as the primary solver indicator, and the relative true residual $\|b - A_{\omega_H}x_k\|_2/\|b\|_2$ as the preconditioner-independent verification quantity on which every stopping tolerance is defined. In the left-preconditioned runs the true residual is recovered by forming $b - A_{\omega_H}x_k$ explicitly, and this cost is accounted for.

For the high-frequency operator A_{ω_H}, the shifted operator has the same discrete Laplacian and a complex-shifted mass term,

$$M_{\mathrm{CSL}} = -\Delta_h - (1 + i\beta)\omega_H^2 I, \tag{6.13}$$

with the same boundary or PML convention as A_{ω_H}. The preconditioner is applied by an exact sparse LU factorisation of M_{CSL}. The shifted operator and its factorisation are assembled once for each $(\omega_H, \text{grid}, \text{PML}, \beta)$ tuple and reused across right-hand sides.

The main two-dimensional runs use damping

$$\beta = 0.3. \tag{6.14}$$

For a warm-start comparison, each test sample is run twice: $x_0 = 0$ and $x_0 = T_{\uparrow,\theta}(u_{\omega_L})$. The right-hand side, high-frequency operator, CSL preconditioner, iteration budget, and stopping rule are identical in both runs. The residual trajectory is recorded at every FGMRES iteration. The main two-dimensional solver evaluation uses 10 test samples per frequency pair and a fixed budget of 40 FGMRES iterations. At $\beta = 0.3$ the reported 2D runs do not reach the target tolerance within this budget, so the convention adopted is to report the final residual after 40 iterations as the main 2D comparison.

6.6.1. Left and right preconditioning

The one-dimensional Dirichlet experiments include an explicit left-versus-right preconditioning comparison. The distinction matters because the residual minimised by FGMRES depends on how the preconditioner is inserted. Left preconditioning solves

$$M^{-1}Ax = M^{-1}b, \tag{6.15}$$

whereas right preconditioning solves

$$AM^{-1}y = b, \qquad x = M^{-1}y. \tag{6.16}$$

The distinction matters for reporting. With right preconditioning, the residual the method minimises is already the true residual $\|b - Ax_k\|_2$, so the relative true residual $\|b - Ax_k\|_2/\|b\|_2$ is read off directly. With left preconditioning, the method instead minimises the preconditioned residual $\|M^{-1}(b - Ax_k)\|_2$;

the true residual is then not the quantity the iteration monitors and must be recovered by forming $b - Ax_k$ explicitly, at the cost of one additional operator application per reported point. In the left-preconditioned comparison the true residual is therefore computed separately and this extra cost is accounted for, while the preconditioned residual $\|M^{-1}(b-Ax_k)\|_2/\|M^{-1}b\|_2$ is reported as the quantity the left-preconditioned Krylov process actually sees.

6.6.2. Fair comparison and cost accounting

Warm-start comparisons must distinguish between two notions of cost. The first is the cost of the high-frequency Krylov solve after the initial guess has been chosen. The second is the total pipeline cost required to produce that initial guess. This distinction is important because the zero start has essentially no setup cost, whereas a learned warm start requires a low-frequency solve, neural-network evaluation, and possible data movement between solver and model representations.

For this reason, the main solver-facing tables report iteration counts and residual histories as *conditional solver costs*: they answer the question whether the high-frequency FGMRES run improves once a particular initial guess is available. In this conditional comparison, the operator A_{ω_H}, the right-hand side b, the CSL preconditioner M_{CSL}^{-1}, the damping parameter β, the stopping rule, and the iteration budget are held fixed. Only x_0 changes. This isolates the algebraic effect of the learned warm start on the high-frequency Krylov process.

A separate end-to-end cost comparison would have to include the setup required to obtain x_0. For a learned warm start this cost can be written schematically as

$$C_{\text{learned}} = C_{\text{low solve}} + C_{\text{model}} + C_{\text{transfer}} + C_{\text{FGMRES}}(x_0), \tag{6.17}$$

whereas the corresponding zero-start cost is

$$C_{\text{zero}} = C_{\text{FGMRES}}(0). \tag{6.18}$$

The learned approach is end-to-end beneficial only if the reduction in high-frequency Krylov work is large enough to compensate for $C_{\text{low solve}}$, C_{model}, and C_{transfer}.

The present thesis primarily reports the conditional solver cost rather than a full wall-clock cost model. This is deliberate. The purpose is to test whether learned frequency transfer supplies algebraically useful information to the high-frequency solve. Wall-clock timings would depend strongly on implementation choices, hardware, batching, sparse-factorisation reuse, and whether the low-frequency solution is already available from a frequency sweep. The iteration and residual comparisons are therefore the fairer first diagnostic. They show whether the warm start changes the Krylov problem in the desired direction before claiming an end-to-end speedup.

Whenever setup costs are relevant to interpretation, they are stated separately. In particular, a reduction in high-frequency FGMRES iterations is not automatically reported as a total speedup. It is reported as evidence that the learned prediction improves the initial algebraic state. A practical speedup claim would require an additional timing study that includes the low-frequency solve, neural-network inference, memory transfer, and any preconditioner setup or reuse.

6.7. Diagnostics and success criteria

The experiments report a hierarchy of diagnostic quantities because no single scalar captures the behaviour of a learned Helmholtz solver component. The hierarchy moves from field prediction to algebraic compatibility: field error measures how well the network predicts the target, while residuals measure whether that prediction is useful to the solver.

M1: field error.

$$E_{\text{field}} = \frac{\|\hat{u}_{\omega_H} - u_{\omega_H}\|_2}{\|u_{\omega_H}\|_2}. \tag{6.19}$$

This is the principal measure of prediction quality. In 2D it is reported on both the full grid and the physical interior; the full-grid value is the primary quantity for the FD/PML models because the solver applies the full-grid operator.

M2: initial true residual.

$$E_{\text{res},0} = \frac{\|b - A_{\omega_H} x_0\|_2}{\|b\|_2}. \tag{6.20}$$

For the zero-start, $x_0 = 0$ and $E_{\text{res},0} = 1$. A learned warm start with $E_{\text{res},0} < 1$ immediately reduces the solver's starting point, but this is not required for the warm start to be useful: a prediction with $E_{\text{res},0} > 1$ may still improve the finite-budget trajectory.

M3: preconditioned residual.

$$E_{\text{prec},0} = \frac{\|M_{\text{CSL}}^{-1}(b - A_{\omega_H} x_0)\|_2}{\|M_{\text{CSL}}^{-1} b\|_2}. \tag{6.21}$$

This is what the Krylov method actually minimises after M_{CSL}^{-1} has acted. It is reported alongside M2 to distinguish true-residual behaviour from preconditioned-residual behaviour.

M4: FGMRES residual curve.

$$k \mapsto \frac{\|b - A_{\omega_H} x_k\|_2}{\|b\|_2}. \tag{6.22}$$

The full curve is the primary solver-facing object. A single iteration count can obscure whether a method benefits from a better starting point, a faster convergence slope, or both.

M5: finite-budget final residual or iteration count. If all methods converge within the fixed budget, the reported cost is the iteration count to tolerance. If none converge, the reported metric is the final residual after the budget. This convention is fixed before inspecting results so that reporting does not depend on the outcome.

Success criteria. The diagnostics are interpreted as a hierarchy rather than as independent scores. Field error first answers whether the network has learned a meaningful approximation of the high-frequency wavefield. The true residual then tests whether that field is compatible with the discrete Helmholtz operator used by the solver. Finally, the FGMRES residual history tests whether any initial algebraic improvement survives the CSL-preconditioned Krylov process.

The success criteria are stage-dependent. For the machine-learning stage, the primary criterion is a small field error $M1$, because this shows whether the network has learned a useful approximation of the high-frequency field. For the solver-facing stage, the primary criterion is whether that approximation reduces the finite-budget true residual or the iteration count relative to the zero start. Because residual improvement depends on the residual-amplification barrier, good field accuracy does not automatically imply solver gain, and both outcomes are reported.

A negative result is informative if the field error is small but the residual benefit is absent. This pattern identifies residual amplification as the limiting factor rather than network quality, and explains the limiting mechanism rather than merely reporting it.

6.8. Verification and sanity checks

Several checks are run before any solver result is interpreted.

Dataset integrity. Each FD/PML dataset is audited for the recorded number of samples, the presence of NaNs or extremely small fields, the stored source-channel convention, and consistency between dataset operator metadata and the evaluator operator.

Source consistency. For each paired sample, the low- and high-frequency fields are generated from the same complex source. Otherwise the learned map would be trained on an ill-defined transfer problem.

Residual convention. The evaluator verifies $r(x) = b - Ax$ and confirms that the zero-start residual satisfies $\left| \|r(0)\|_2 / \|b\|_2 - 1 \right| < 10^{-14}$, a tolerance check rather than an exact floating-point equality. This catches sign errors, scaling mistakes, and mismatched source normalisation.

PML compatibility. For PML experiments the full-grid residual is always inspected alongside interior field error. A model that performs well in the interior but produces large PML energy or a large full-grid residual is not considered solver-compatible.

One-dimensional spectral diagnostic. In the Dirichlet diagnostic problem, the modal identity $c_k(r_0) = \lambda_k\, c_k(e_0)$ is checked numerically per sample through the relative residual

$$\delta_{\text{modal}} \;=\; \frac{\|c(r_0) - \lambda \cdot c(e_0)\|_2}{\|c(r_0)\|_2 + \varepsilon}, \qquad \varepsilon = 10^{-14}, \tag{6.23}$$

where $c(r_0)$ and $c(e_0)$ are the Dirichlet modal coefficients of the initial residual and initial error. This quantity is reported as a diagnostic in each run rather than gated by a pass/fail tolerance.

Sequential-split balance. The 1D pipelines use a sequential train/validation split rather than a random permutation. To confirm that this convention does not introduce systematic bias, the per-channel means and standard deviations of $|u_{\omega_L}|$, $|u_{\omega_H}|$, and ρ are compared between the training and validation blocks. Across the four 1D datasets the relative difference between train and validation block means is at most $5.1\,\%$ for $|u_{\omega_H}|$, at most $1.8\,\%$ for $|u_{\omega_L}|$, and at most $1.5\,\%$ for ρ. The split is therefore treated as benign.

Network learning sanity and overfitting checks. Two separate checks are distinguished. First, small-subset training runs verify that the implementation can reduce the supervised field loss when the optimisation problem is deliberately made easy. This is a training-loop sanity check, not evidence that the model predicts accurately on unseen samples drawn from the same distribution. Second, the production runs are monitored through the train–validation gap. When validation loss bottoms out while training loss continues to decrease, later epochs are treated as overfitting and checkpoint selection uses the best validation epoch rather than the final epoch. Solver-facing claims are evaluated only on held-out validation or test samples.

All training data were screened before use. Samples containing a `NaN` or `Inf` entry in either field component were discarded. Samples whose interior RMS fell below 10^{-8}, or whose interior RMS after normalisation differed from unity by more than 0.05, were also rejected. The full screening thresholds are reported in Table 10.2 in Appendix 10.

For the main two-dimensional warm-start experiments the CSL-preconditioned FGMRES solver was run for a fixed budget of 40 iterations. Older debugging runs with different budgets are treated as exploratory and are not used for the main solver-facing comparison.

6.9. Computational workflow, artefacts, and safeguards

The computational workflow has four stages: dataset generation, neural-network training, solver evaluation, and diagnostic analysis. Dataset generation and solver evaluation are dominated by sparse Helmholtz solves; neural-network training is dominated by GPU throughput; one-dimensional spectral diagnostics are small enough to support dense checks or analytical eigenbasis calculations.

Experiment outputs are kept in named directories that record the frequency pair, boundary or PML convention, training objective, and diagnostic family. Corrected pipelines are stored separately from exploratory pipelines rather than overwriting them. This matters because several conventions can produce similar-looking wavefield plots while representing different algebraic systems: Green-function data, interior-focused PML data, row-scaled PML data, and corrected full-grid FD/PML data should not be mixed without explicit labels.

For thesis-quality runs the reproducibility record consists of the launch script or sweep specification, numerical metadata, the trained checkpoint, a summary table, the residual curves, and the figure-generation script. Training checkpoints such as `best.pt` and `last.pt` are retained where applicable, so that a model can be evaluated or resumed from its saved checkpoints without rerunning the job from the start. Derived quantities such as residuals, modal coefficients, and plots are treated as regenerable analysis artefacts rather than primary dataset fields.

The hardware organisation follows from these stages but is not itself a scientific variable. Local runs are used for syntax checks, small validation tests, and plotting; interactive GPU servers are used for training and rapid solver debugging; cluster resources are used for long sweeps and repeated benchmarks. Across all environments the rule is to use the smallest computation that answers the current question reliably and to reserve large runs for designs that have already passed small-scale validation. The hardware environments themselves are recorded in Appendix 10.

Table 6.5: Verification and sanity checks used before interpreting solver-facing results. These checks rule out implementation, data, and evaluation errors as explanations for the observed behaviour.

Component	Check	Consequence
Dataset integrity	NaN/Inf screening, nonzero interior RMS, metadata consistency	Training pairs are treated as valid Helmholtz transfer samples.
Source consistency	Same complex source used for u_{ω_L} and u_{ω_H}	The supervised map is a well-defined frequency-transfer task.
Residual convention	zero-start residual satisfies $\left\lvert \lVert b - A0\rVert/\lVert b\rVert - 1 \right\rvert < 10^{-14}$	Residual curves are comparable across starts.
FD/PML compatibility	Full-grid residual inspected alongside interior field error	Interior accuracy alone is not accepted as solver compatibility.
1D modal diagnostic	Numerical check that $c_k(r_0) \approx \lambda_k c_k(e_0)$ up to rounding	The residual-amplification analysis uses the same operator as the solve.
Train/validation balance	Distributional comparison of split statistics	Sequential or random splits are not treated as the source of the effects.
Network learning sanity	Small-subset training and train–validation gap	Implementation failure is separated from overfitting and from performance on unseen data.

Table 6.6: Computational stages used in the experiments. The final evaluation is solver-native: field loss is recorded, but residuals and FGMRES behaviour decide whether a variant is useful.

Stage	Main bottleneck	Primary outputs
Dataset generation	sparse solves and I/O	fields, sources, metadata
Neural-network training	GPU throughput	checkpoints, validation losses
Solver evaluation	sparse solves and FGMRES	residuals, iterations, curves
Diagnostic analysis	spectra, plots, summaries	modal and PML diagnostics, figures

Table 6.7: Computational risks and the safeguards adopted in the final experiments.

Risk	Consequence	Safeguard
Wrong discretisation convention	Misleading residuals or spectra	Validate on the 1D case first
Field loss improves only visually	No Krylov benefit	Evaluate true residuals and FGMRES curves
PML incompatibility	Inflated full-grid residual	Train and evaluate full-grid FD/PML fields
Untraceable sweeps	Irreproducible comparisons	Named output folders and summary files
Repeated setup costs	Distorted timing interpretation	Reuse sparse operators and factorisations

6.10. Chapter summary

This chapter defined the experimental design and computational setup used to test learned frequency transfer as a Helmholtz solver component. The design separates three questions that are easy to conflate. The first is whether a neural model can predict a high-frequency field from a low-frequency field. The second is whether the predicted field has a small residual under the high-frequency operator. The third is whether the resulting initial state or learned correction reduces FGMRES work.

The three questions are answered by different evaluations on three settings. The one-dimensional Dirichlet problem provides an exact spectral microscope through the modal identity $c_k(r_0) = \lambda_k\, c_k(e_0)$. The one-dimensional PML problem isolates absorbing-layer compatibility without the cost and complexity of the two-dimensional setting. The two-dimensional FD/PML problem is the solver-facing evaluation, with full-grid training targets and full-grid residuals against the operator that FGMRES applies. The metric hierarchy, success criteria, and verification checks are stated in advance so that the Results chapter can evaluate both positive and negative outcomes against the same standard.

Part III

Results and Conclusions

7

Results

This chapter evaluates whether learned frequency transfer produces solver-useful warm starts for CSL-preconditioned FGMRES applied to the Helmholtz equation. The central tension throughout is that a learned model is trained to minimise field error, while the Krylov solver is driven by residuals. These two objectives are related through the identity $r_0 = A_{\omega_H} e_0$, but they are not interchangeable. A field-accurate prediction can still produce an elevated initial residual if its error concentrates in modes that the discrete operator amplifies strongly. Every section of the chapter is organised around this distinction.

Metric hierarchy used in this chapter. Three quantities are reported throughout, and they answer different questions. *Field error* $\|x - u_\star\|_2/\|u_\star\|_2$ is the scientific target. It measures how close the iterate is to the true solution and is the quantity the learned model is trained against. The *preconditioned residual* $\|M_{\text{CSL}}^{-1}(b-Ax)\|_2/\|M_{\text{CSL}}^{-1}b\|_2$ is the primary solver indicator. It is the quantity the CSL-preconditioned Krylov process actually minimises, and it tracks the field error far more faithfully than the true residual does, as Section 7.4 demonstrates directly. The *true residual* $\|b - Ax\|_2/\|b\|_2$ is reported for verification, but it is sensitive to mode-by-mode amplification by the indefinite operator and can be large even when the iterate is close to the exact solution. When the three metrics disagree, the disagreement is itself a result, not a metric problem.

The chapter proceeds as a chain of evidence in three stages. The first stage (Section 7.2) evaluates the learned model on its own terms as a field predictor and establishes what physical structure the network has learned. The second stage (Sections 7.4 and 7.5) isolates the residual-amplification mechanism in one dimension, where the eigenbasis is explicit and the mechanism can be diagnosed exactly. The third stage (Section 7.6) tests whether the one-dimensional lessons survive in the full two-dimensional FD/PML setting. Section 7.8 collects the conclusions.

All solver-facing results use CSL damping $\beta = 0.3$. Throughout the chapter, the learned model acts only as a warm start. It sets the initial guess x_0 and therefore changes the initial residual, but it does not modify the Helmholtz operator or the CSL preconditioner. Each result is interpreted only at the level it directly supports. Field-error results are used to assess learned frequency transfer, residual results are used to assess algebraic compatibility, and FGMRES curves are used to assess solver behaviour. A positive result in one category is not automatically treated as a positive result in the others.

7.1. Conventions and method labels

This section fixes the objects, metrics, and initial-guess labels used throughout the chapter, so that the result sections can refer to them without redefinition.

7.1.1. The learned operator

The learned operator maps a low-frequency solution field to its high-frequency counterpart at the next frequency pair. The input channels are the real and imaginary parts of the low-frequency field. Right-hand-side-aware variants add the real and imaginary source channels; full channel specifications, including the optional damping profile $\sigma(x)$ introduced in Section 3.7, are given in Section 6.5. The output is the predicted high-frequency field.

A one-dimensional U-Net is used for the Dirichlet and 1D PML studies. In the two-dimensional FD/PML problem the main model is a full-grid `TransferUNet` with `base_ch=32` and five encoder and decoder levels. Its output is defined on the full 512×512 computational grid, including the PML region. A separate analytic Green's-function CNN, referred to throughout as the *analytic diagnostic CNN*, is used only for the interpretability diagnostic in Section 7.3; it is not part of the FD/PML solver benchmark. Its run label and checkpoint path are recorded in Appendix 10.

7.1.2. Field-prediction error

Field quality is measured before the predicted field is inserted into the solver. The metric is the complex relative ℓ^2 error,

$$\mathrm{RelL2}(x) = \frac{\|x - u_\star\|_2}{\|u_\star\|_2 + \varepsilon}, \tag{7.1}$$

where $u_\star$ is the reference high-frequency solution, the norm acts on the complex grid vector, and $\varepsilon > 0$ is a small stabilisation constant added once here as a division-by-zero guard. A value of 0 is a perfect prediction; the zero initial guess has RelL2 $= 1$ by construction. In the FD/PML setting the norm can be evaluated on the full 512×512 grid (including the PML) or on the physical 288×288 interior only. Both are reported to separate model accuracy in the physical domain from accuracy in the absorbing layer; the PML region is systematically harder to predict. The full-grid versus interior-only comparison also prepares the solver experiments in Section 7.5, which test whether operator compatibility requires predicting the full grid.

7.1.3. Initial-guess constructions

Every solver table in this chapter compares different choices of initial guess x_0. The Helmholtz operator, CSL damping, preconditioner, and Krylov algorithm are fixed within each experiment. Only the initial state changes. Table 7.1 defines the labels used throughout.

Table 7.1: Initial-guess constructions compared in this chapter. All learned starts set x_0 only; none modifies the Helmholtz operator or the CSL preconditioner.

Label	Role	Construction
Zero start	Baseline	$x_0 = 0$; unit field error and unit residuals.
Raw learned start	Solver test	Full network output used directly as x_0.
Interior-only learned start	Injection test	Physical-interior prediction kept; PML strip zeroed before the solve.
Residual-gated start	1D diagnostic	Modal components kept only if they reduce the modal residual.
Analytic-PML raw	Mechanistic diagnostic	Analytic Green's-function model output used directly.
Analytic-PML interior-only	Mechanistic diagnostic	Same model, PML strip zeroed.
Interior-trained depth-five	Model variant	Five-level 2D model trained on interior targets only.
Full-grid FD/PML-trained	Main model	`TransferUNet` trained on full FD/PML targets including the PML.

The interior-only construction tests whether the learned physical-domain field is useful when the predicted PML is not trusted. It is a controlled diagnostic injection choice. Zeroing the PML strip introduces a mismatch at the interior to PML interface, so a worse residual for the interior-only start does not imply that the interior prediction is useless.

7.2. Machine-learning results

The architectural requirements for learned frequency transfer were derived in Section 5.8, and the specific models, input channels, and training objectives are specified in Section 6.5. This section reports how the trained models behave: the field-prediction errors they achieve, the stability of their training, and the physical structure their predictions encode. Three questions are separated: does the network learn a meaningful cross-frequency map; where is that map accurate and where does it fail; and what physical structure does it encode. The solver-facing consequences follow in Sections 7.4 through 7.6.

7.2.1. Field errors across frequency pairs

The learning task is frequency transfer. Given u_{ω_L}, predict u_{ω_H} at the next frequency pair. The tested pairs are $16 \to 32$, $32 \to 64$, and $64 \to 128$. Table 7.2 reports the best validation RelL2 of the full-grid FD/PML-trained models.

Table 7.2: Best validation relative ℓ^2 field error of the full-grid FD/PML-trained models. Interior errors are lower than full-grid errors, confirming that the physical domain is easier to predict than the complete FD/PML solution including the absorbing layer.

Pair	Full-grid RelL2	Interior RelL2
$16 \to 32$	0.240	0.190
$32 \to 64$	0.321	0.252
$64 \to 128$	0.339	0.276

The network reduces full-grid field error from 1 to 0.24 through 0.34 across all pairs. This supports the learning claim: the low-frequency field contains usable information about the high-frequency field, and the network extracts a nontrivial part of that structure. The result should not yet be read as a solver claim. The metric in Table 7.2 is a field-prediction metric, not a residual metric and not an iteration-count metric.

The frequency trend is also informative. Field error grows with the target frequency, which is expected. At $\omega_H = 128$ the field contains more oscillations across the same physical domain, so small phase errors produce larger relative errors. The gap between full-grid and interior errors is consistent across pairs and confirms that the PML region is harder to predict than the physical interior, while not so difficult that full-grid training becomes ineffective.

At the same time, the errors are not small enough to treat the network as a direct high-frequency solver. Interior error of 0.19 to 0.28 represents a structured approximation, not a replacement for the FD/PML solve. Whether this approximation is useful to CSL-FGMRES depends on where the remaining error is located relative to the high-frequency operator, which is tested only in the solver-facing sections.

7.2.2. Role of the PML in the learned prediction

The consistently lower interior error confirms that the network predicts the physical solution better than the complete FD/PML field. This does not justify ignoring the PML in solver experiments. The PML is part of the discrete operator. A prediction that is accurate in the interior but inconsistent in the absorbing layer or at the interface can still produce a large residual when the full FD/PML operator is applied. Section 7.5 isolates this effect quantitatively. The raw versus interior-only comparison in all subsequent tables is the direct test of PML operator compatibility.

7.2.3. Training behaviour

Figure 7.1 shows that training and validation losses decrease stably for the hardest pair. This supports the claim that the reported field errors arise from a learned frequency-transfer map rather than from a failed or unstable optimisation process. The figure does not, by itself, establish solver usefulness. It validates the supervised field-prediction stage. Residuals and FGMRES curves are needed to evaluate the warm start as a solver component.

7.3. Emergent source-wise decomposition

The field-error values establish that the network has learned a useful cross-frequency map, but the loss numbers do not reveal what that map encodes internally. The internal structure has direct solver consequences: error components are amplified into residuals by their eigenvalue magnitude $|\lambda_k|$ (Proposition 5.1), so a model that concentrates its errors in operator-sensitive directions produces a larger initial residual than its field error alone suggests. The question pursued here is therefore precise. Has the network learned to decompose a multi-source wavefield into individual source contributions, even though it was never given source labels, source counts, or any structural supervision that would reward such a decomposition?

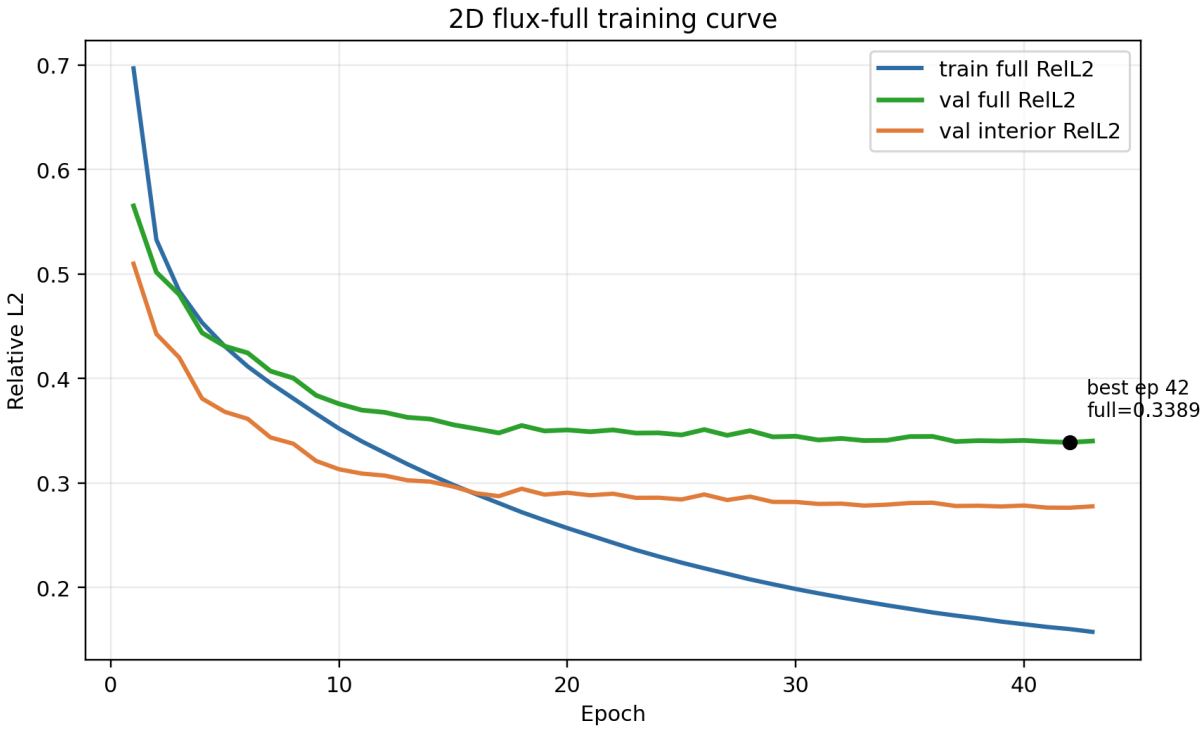


Figure 7.1: Training and validation curves for the FD/PML `TransferUNet` on the $64 \to 128$ pair. The full-grid validation curve measures RelL2 over the complete grid including the PML; the interior curve applies the same metric to the physical domain only. Both curves decrease stably, confirming that the model has learned a genuine cross-frequency map at the hardest tested pair. The persistent gap between curves quantifies the additional difficulty of predicting the absorbing layer.

This section reports a positive answer, together with the diagnostic that rules out the most plausible competing explanation.

The diagnostic is conducted on a CNN trained exclusively on analytic free-space multi-source fields (the analytic diagnostic CNN, Section 7.2.1); the FD/PML `TransferUNet` used in the solver experiments is trained on single-source fields and is therefore not amenable to source-wise decomposition analysis by construction. The analytic diagnostic CNN uses 29 input channels: the real and imaginary parts of the normalised low-frequency field, 24 Fourier coordinate features, a static PML damping map built from the profile $\sigma(x)$ (see Section 3.7 for the definition of σ and its role in the absorbing boundary), a frequency channel, and a zero-valued placeholder channel. No source-location information is supplied to the network at training or inference time.

Mathematical setup. For a Helmholtz problem driven by multiple sources, the solution is a coherent superposition of single-source response components,

$$u_\omega(x) = \sum_i a_i \, \phi_\omega(x;\, x_i), \tag{7.2}$$

where $\phi_\omega(\cdot;\, x_i)$ is the free-space Green's function centred at source x_i and $a_i \in \mathbb{C}$ is its complex amplitude. Each term $a_i \phi_\omega(x;\, x_i)$ is a *single-source response component*: the field produced by source i acting alone. The network receives only the superposed low-frequency field; the individual components, source locations, and amplitudes are withheld entirely.

The Voronoi tessellation of the source locations provides a natural interpretability foil. It assigns each spatial point to its nearest source, and the corresponding Voronoi-windowed approximation is

$$\widehat{u}^{\mathrm{Vor}}_{\omega_H}(x) = \sum_i W_i(x)\, a_i \, \phi_{\omega_H}(x;\, x_i), \qquad W_i(x) = \mathbf{1}[i = \arg\min_j \|x - x_j\|_2]\,. \tag{7.3}$$

Because the indicators $\{W_i\}$ form a partition of unity ($\sum_i W_i(x) = 1$ almost everywhere), the approximation covers the entire domain without gaps; every spatial point belongs to exactly one cell. The Voronoi construction is therefore the best prediction a purely local nearest-source rule could achieve. It is used not as a model but as a diagnostic. By comparing the network prediction and its errors against this construction, the analysis separates what local source-wise propagation can explain from what requires global knowledge of the interference between sources.

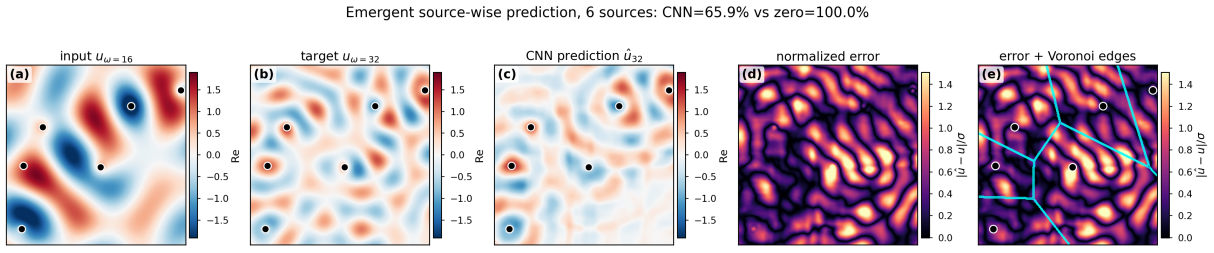


Figure 7.2: Emergent source-wise structure in the analytic CNN ($16 \to 32$, six sources). Panel (a): real part of the input low-frequency field u_{ω_L}. Panel (b): real part of the target high-frequency field u_{ω_H}. Panel (c): real part of the CNN prediction $\widehat{u}_{\omega_H}$. Panel (d): normalised prediction error $|\widehat{u}_{\omega_H} - u_{\omega_H}|/\|u_{\omega_H}\|_\infty$. Panel (e): same error with Voronoi cell boundaries (cyan) and source locations (white dots) overlaid. The prediction preserves outgoing wave packets centred on each source; the error does not concentrate exclusively at Voronoi boundaries, indicating that cross-cell interference contributes to the remaining error.

The network preserves source-centred wave structure. Figure 7.2 shows a representative six-source example. The network prediction recovers the outgoing wave packets centred on each source. The Voronoi partition, computed *after* inference from source locations the network never saw, aligns approximately with regions of elevated error. The error is not confined to cell boundaries, however. It spreads across the domain, which indicates that something beyond nearest-source propagation is missing.

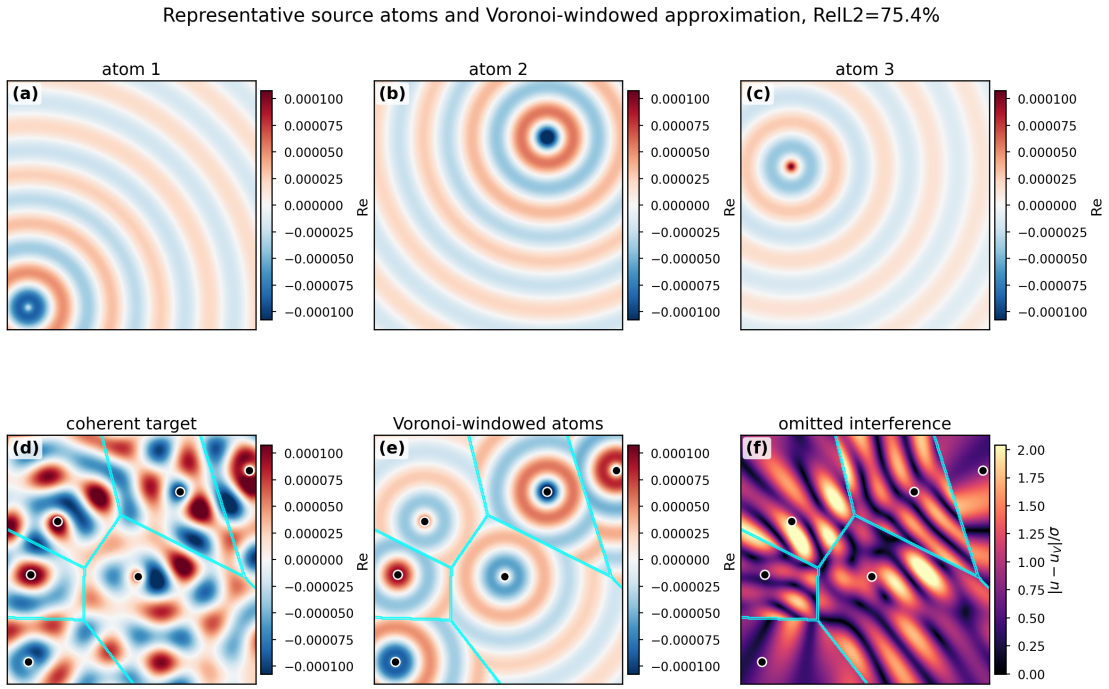


Figure 7.3: Decomposition diagnostic for the same six-source example. Panels (a) through (c): real parts of three representative single-source response components $a_i \phi_{\omega_H}(x; x_i)$ at the target frequency, shown at the scale of the full superposed field. Panel (d): real part of the coherent target u_{ω_H}, the phase-sensitive sum of all six components. Panel (e): real part of the explicitly constructed Voronoi-windowed approximation, Equation (7.3), the best achievable by a purely local nearest-source rule. Panel (f): omitted interference $\delta(x)$, Equation (7.4), the pointwise magnitude of what the Voronoi approximation cannot represent. The omitted interference is concentrated near cell boundaries where contributions from two or more sources overlap.

Local source-wise propagation is insufficient. Figure 7.3 makes the limitation explicit. The Voronoi assembly (panel e) captures the dominant source in each region, but the coherent target (panel d) contains phase-sensitive cancellations and reinforcements that require simultaneous knowledge of multiple sources. Define the *omitted interference* $\delta(x)$ (also referred to as the omitted-component error) at pixel x as

$$\delta(x) = \left| u_{\omega_H}(x) - \sum_i W_i(x)\, a_i \phi_{\omega_H}(x;\, x_i) \right|, \tag{7.4}$$

normalised by the local field amplitude. This is the portion of the high-frequency field that no local source-wise model can reproduce.

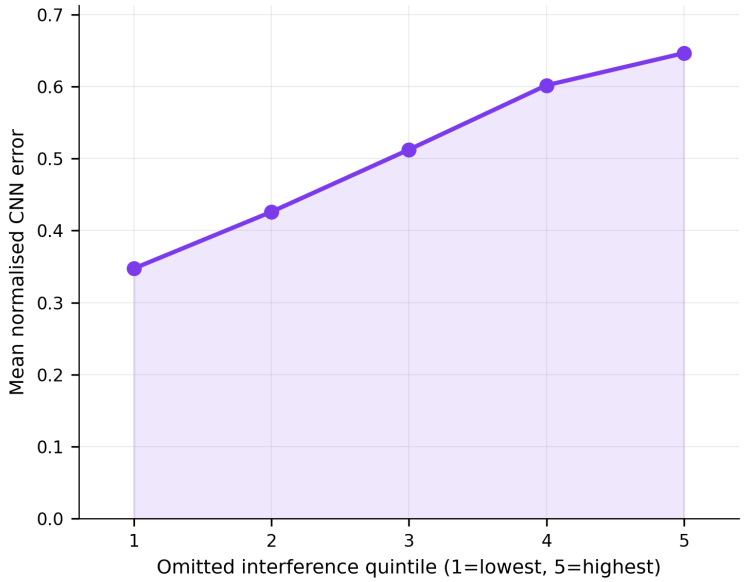


Figure 7.4: Mean normalised CNN prediction error versus omitted-interference magnitude, binned into five equal-population quintiles across 96 analytic test samples ($16 \to 32$, analytic diagnostic CNN, $N = 600$ checkpoint). Error increases monotonically from 0.347 in the lowest quintile to 0.646 in the highest, a ratio of 1.89 (median 1.92 across samples). The monotone rise confirms that the CNN's errors are concentrated precisely where local source-wise propagation is insufficient.

Omitted interference predicts CNN error. Figure 7.4 quantifies the relationship. Pixels are grouped into five equal-population quintile bins by $\delta(x)$; the mean normalised CNN error rises monotonically from 0.347 (lowest quintile) to 0.646 (highest), a ratio of 1.89. The CNN's residual errors are not distributed uniformly. They concentrate in the spatial regions where the omitted interference is large. The network has therefore learned what a local source-wise model captures, and fails where global multi-source interference is required.

Ruling out the Green's-function trap. The structural evidence above could, in principle, be a side-effect of the network passively reproducing the Hankel kernel $G_\omega(x, x_i) = \frac{i}{4} H_0^{(1)}(\omega \|x - x_i\|)$ pointwise. The kernel has a localised amplitude peak at each source, so a network that simply reproduced and superposed G_ω would exhibit apparent windowing purely as a consequence of those peaks, having learned no decomposition at all. This failure mode is referred to here as the *Green's-function trap*; it is a concrete instance of shortcut learning, in which a network latches onto an incidental statistical cue rather than the intended structure [34].

The amplitude-ablation experiment closes this hypothesis directly. The analytic training set is regenerated with the geometric-spreading factor $1/\sqrt{r}$ removed from the kernel, so that each single-source contribution becomes a phase-only wave $e^{i\omega r}$ with no localised amplitude peak. All other settings (architecture, optimiser, sample count, frequency pair, channel layout) are held fixed. The only change is the training data.

The result is a complete collapse of prediction quality. Validation RelL2 is stuck at approximately 1.00 from the first epoch onwards in the amplitude-ablation run, whose checkpoint and run label are recorded in Appendix 10. The network reduces to the zero predictor and does not improve over training. A network operating under the Green's-function trap would degrade gracefully under this ablation, since phase information alone would still permit pointwise reproduction up to amplitude calibration. The observed total collapse is inconsistent with that interpretation. The amplitude singularity is therefore not an incidental feature of the training data. It is the structural cue the network keys on to localise sources. The Voronoi-aligned source-wise structure is consequently not reducible to passive Green's function reproduction.

Implication for solver usefulness. The omitted interference is a globally coherent, spatially extended feature of the high-frequency field. Globally coherent modes at frequency ω_H tend to carry large $|\lambda_k|$, and the residual-amplification identity $c_k(r_0) = \lambda_k c_k(e_0)$ (Section 5.5) then converts these missed interference terms into disproportionately large residual contributions. The Voronoi diagnostic therefore explains qualitatively why a field-accurate prediction can still produce an elevated initial residual. The remaining error concentrates precisely in the modes the operator amplifies most strongly. The 1D Dirichlet experiment in Section 7.4 confirms this prediction quantitatively with explicit eigenmode decompositions.

This interpretability result should be read as a mechanism study, not as direct evidence for the FD/PML solver benchmark. It is obtained with a separate analytic multi-source model, whereas the solver-facing FD/PML model is evaluated through residuals and FGMRES curves. Its value is that it identifies a plausible error structure, namely missed global interference, that explains why good field prediction can still fail to be residual-safe.

7.3.1. Conclusion of the machine-learning results

Three findings emerge from the ML evaluation. First, frequency transfer is learnable. The network reduces full-grid RelL2 from 1 to 0.24 through 0.34 across all tested pairs, with stable generalisation confirmed by training curves. Second, the PML region is harder to predict than the physical interior, but the network cannot safely ignore it because the residual is computed by the full FD/PML operator. Third, and most strikingly, the network spontaneously decomposes multi-source wavefields into source-centred outgoing contributions without any source supervision. Its errors concentrate in the globally coherent interference terms that local models cannot represent, and those same terms carry the greatest residual risk under the high-frequency Helmholtz operator.

These findings set the agenda for the solver experiments. The network has learned a physically meaningful but incomplete map. It captures individual source propagation well and fails on global interference. Whether that incompleteness is fatal to solver usefulness, or merely a limitation that the Krylov process absorbs after a brief transient, is the question the following sections answer.

7.4. One-dimensional Dirichlet results

The one-dimensional Dirichlet experiment isolates the residual-amplification mechanism in the cleanest possible setting: a real symmetric operator with an explicit eigenbasis. The problem is the homogeneous Dirichlet Helmholtz equation on a grid with $N = 512$ unknowns and no PML, with pair $\omega_L = 16$, $\omega_H = 32$. The operator $A_\omega = -D_{xx} - \omega^2 I$ has the real orthonormal eigenpairs (λ_k, v_k) given in Equation (6.5) of Chapter 6, and it is exactly that explicit eigenbasis that makes the modal identity below an equality rather than a bound. The learned model is a 1D U-Net trained on multi-source analytic data to map u_{16} to u_{32}. Its best validation RelL2 is 4.34×10^{-3}, so it is a strong field predictor in this controlled setting.

7.4.1. Field accuracy is not residual accuracy

Proposition 5.1 established the one fact this section relies on: in the Dirichlet eigenbasis the high-frequency operator does not treat all error directions equally, but multiplies the field-error coefficient in each eigenmode by that mode's eigenvalue before the error reaches the residual. The initial field error $e_0 = u_\star - x_0$ and residual $r_0 = A_{\omega_H} e_0$ are therefore related mode by mode by

$$c_k(r_0) = \lambda_k \, c_k(e_0), \tag{7.5}$$

so a small field error in a mode with large $|\lambda_k|$ becomes a large residual coefficient. This is the central mechanism behind the behaviour of learned warm starts, and the Dirichlet setting makes it exact.

Table 7.3: One-dimensional Dirichlet initial-state diagnostics, $16 \to 32$, $\beta = 0.3$. The raw learned start is field-accurate but produces a true residual $11.6\times$ the right-hand-side norm. Residual gating recovers operator compatibility while preserving most of the field improvement. Rows are aligned to the same frozen checkpoint.

Initial guess	Field error	$\|r_0\|_2/\|b\|_2$	$\|M_{\mathsf{CSL}}^{-1} r_0\|_2/\|M_{\mathsf{CSL}}^{-1} b\|_2$
Zero start	1.000	1.000	1.000
Raw learned start	0.084	11.64	0.181
Residual-gated start	0.075	0.740	0.151
Two gated cycles	0.060	0.681	0.129

Table 7.3 is a direct empirical confirmation of Proposition 5.1. The raw learned start achieves field error 0.084, a $12\times$ improvement over the zero start. Its true initial residual, however, is $11.64\|b\|_2$, far above the zero-start value of 1. This is not a failure of the Krylov implementation and not a numerical artefact. It is Equation (7.5) at work. The learned prediction places part of its remaining field error in modes with large $|\lambda_k|$, and the operator converts those components into a large residual.

The same initial state looks much less severe in the preconditioned residual: 0.181, an 80% reduction relative to the zero start. This is the metric hierarchy of the chapter introduction in action. Field error is the scientific target and the learned start improves it substantially; the preconditioned residual is the primary solver indicator and the learned start improves it as well; the true residual is large for this start, but this is exactly the kind of spectral artefact the preconditioner exists to absorb. Interpreting the true residual as a verdict on the learned start would be reading the wrong metric.

The residual-gated start confirms that the learned field contains genuinely useful solver-facing information. The gate accepts a modal component only if it reduces the modal residual relative to the current baseline. It requires the exact Dirichlet eigenbasis and is therefore a diagnostic tool, not a production method. After gating, the true residual falls to 0.740, below the zero-start value, while field error also improves slightly to 0.075. The learned prediction is not intrinsically unsafe. Its harmful and helpful components are mixed together in the raw output.

Figures 7.5 and 7.6 make the transient nature of the residual elevation visible. After the early iterations, all curves settle into the same CSL-controlled convergence slope. The learned start does not alter the preconditioned operator, only the initial state.

7.4.2. Left and right preconditioning

Left preconditioning ($M^{-1}Ax = M^{-1}b$) minimises the preconditioned residual directly. Right preconditioning ($AM^{-1}y = b$, $x = M^{-1}y$) more directly targets the true residual of the original system. The preconditioned residual is closer to the field error than the true residual is, so left preconditioning is also the choice most aligned with the metric the learned model is trained against. The comparison below verifies that this expectation holds empirically in the controlled 1D setting.

A terminological point avoids a later ambiguity. Flexible preconditioning is defined on the right, so the left-preconditioned runs reported here are standard GMRES applied to the left-preconditioned system $M_{\mathsf{CSL}}^{-1}Ax = M_{\mathsf{CSL}}^{-1}b$; the "FGMRES" label is retained only because the same solver harness is used throughout with its flexibility disabled. This is legitimate because M_{CSL} is a fixed sparse LU factorisation, so the preconditioner does not vary between iterations and the flexible and non-flexible iterations coincide.

The two sides minimise different quantities, so neither dominates uniformly. Right preconditioning produces smaller iter-10 true residuals, left preconditioning produces smaller iter-10 preconditioned residuals. Under the metric hierarchy stated in the chapter introduction, namely preconditioned residual as the primary solver indicator and true residual as verification, left preconditioning is the appropriate choice. All subsequent experiments in this chapter therefore use left preconditioning.

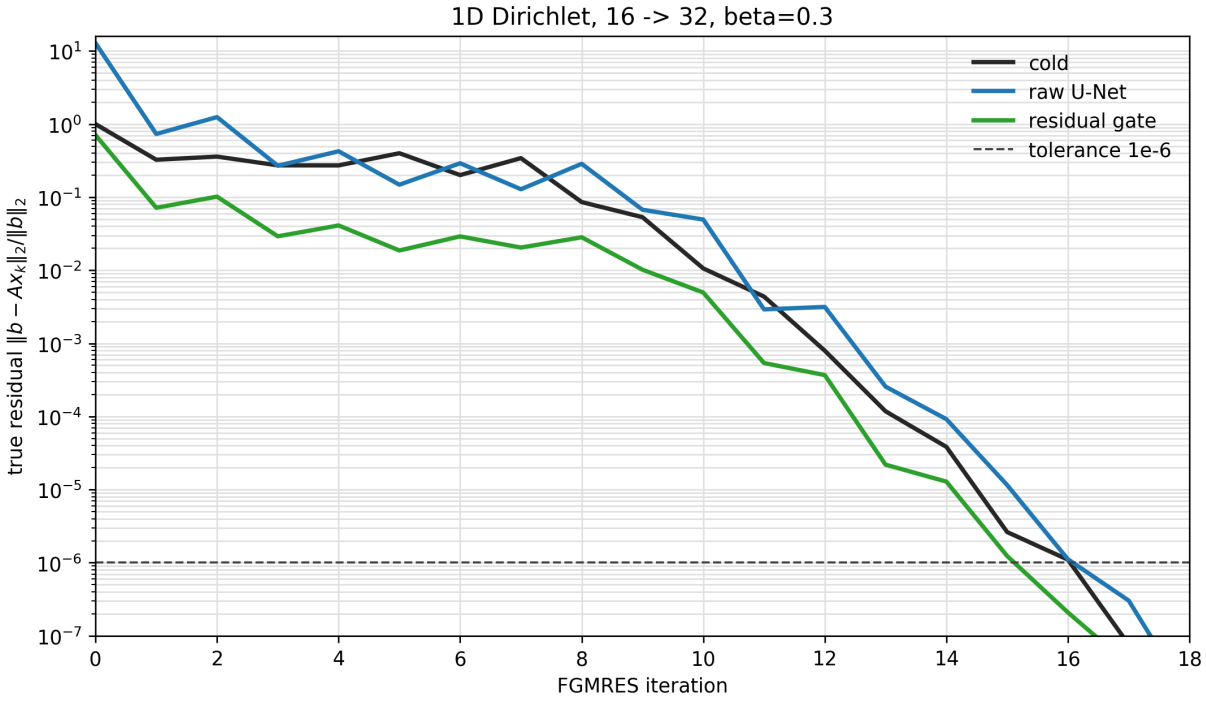


Figure 7.5: True residual versus FGMRES iteration, 1D Dirichlet $16 \to 32$, $\beta = 0.3$. The raw learned start enters with residual $11.6\times$ the zero start but converges to tolerance at the same iteration count: the elevated entry is a transient, not a persistent penalty. The residual-gated start begins below the zero-start residual and converges first.

Table 7.4: Left versus right CSL preconditioning, 1D Dirichlet $16 \to 32$. Initial residuals are identical because the initial guess is the same; the divergence at iteration 10 reflects the different quantities each side minimises.

Method	Side	Initial true res.	Iter.-10 true res.	Iter.-10 prec. res.
Zero start	Left	1.000	1.07×10^{-2}	2.74×10^{-3}
Zero start	Right	1.000	2.75×10^{-3}	4.45×10^{-3}
Raw learned start	Left	11.64	3.87×10^{-2}	6.39×10^{-4}
Raw learned start	Right	11.64	6.37×10^{-4}	1.42×10^{-3}
Residual-gated start	Left	0.740	4.31×10^{-3}	5.31×10^{-4}
Residual-gated start	Right	0.740	5.56×10^{-4}	1.16×10^{-3}

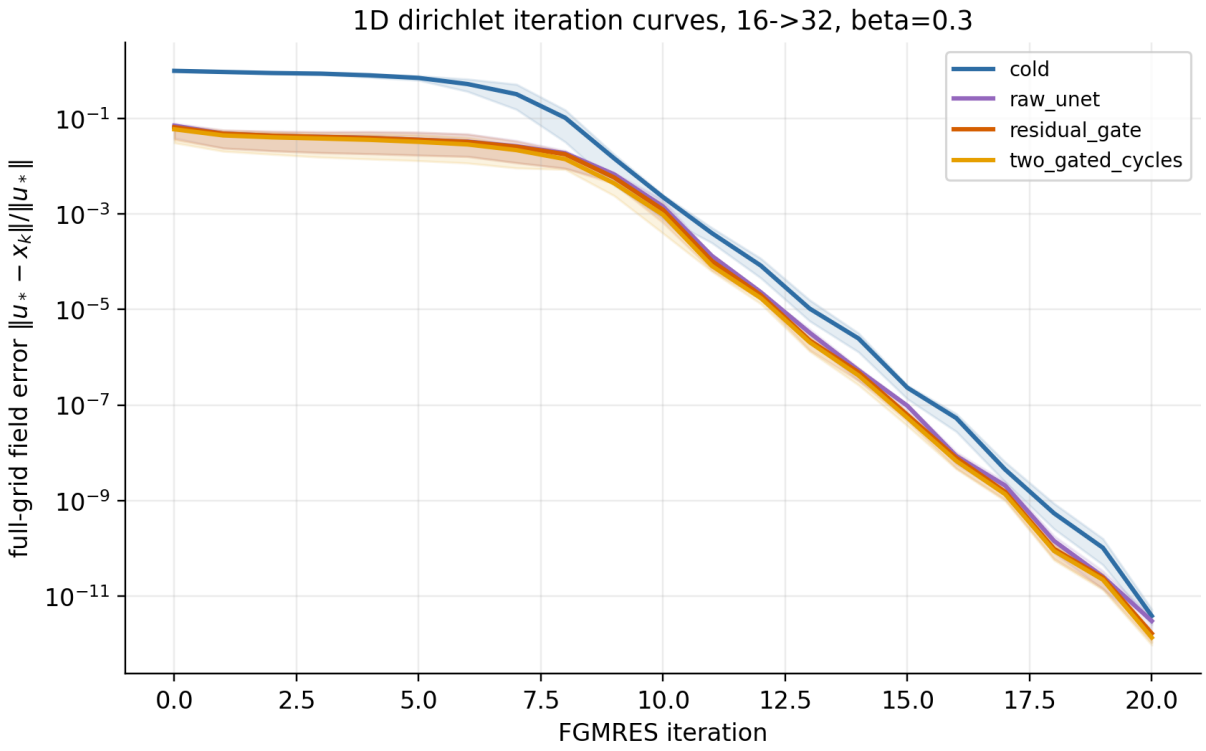


Figure 7.6: Field error versus FGMRES iteration for the same experiment. The learned starts are much closer to the exact solution at iteration zero. After the Krylov process repairs the initial residual mismatch, all curves approach the same convergence slope. Field error and true residual must be read together. They start from different algebraic states and rejoin only once the CSL-preconditioned operator dominates.

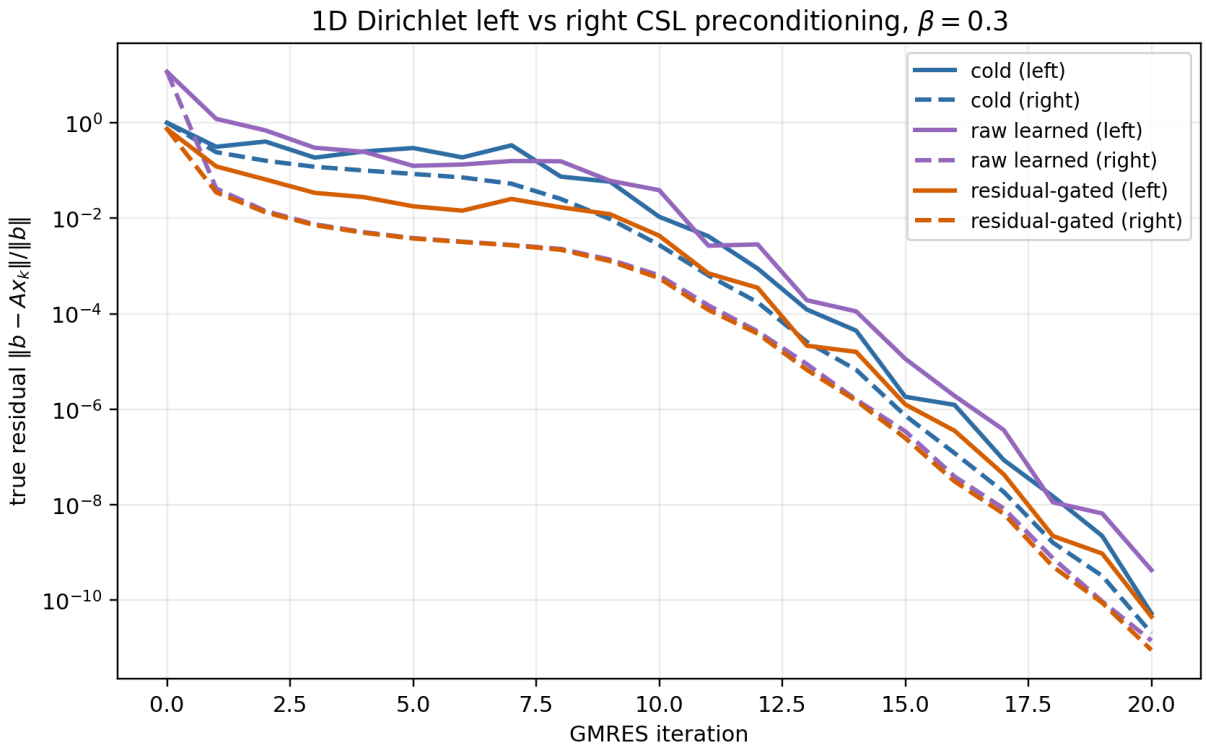


Figure 7.7: True residual, left- versus right-preconditioned FGMRES, 1D Dirichlet. Right preconditioning reduces the true residual more directly because it minimises over the original system.

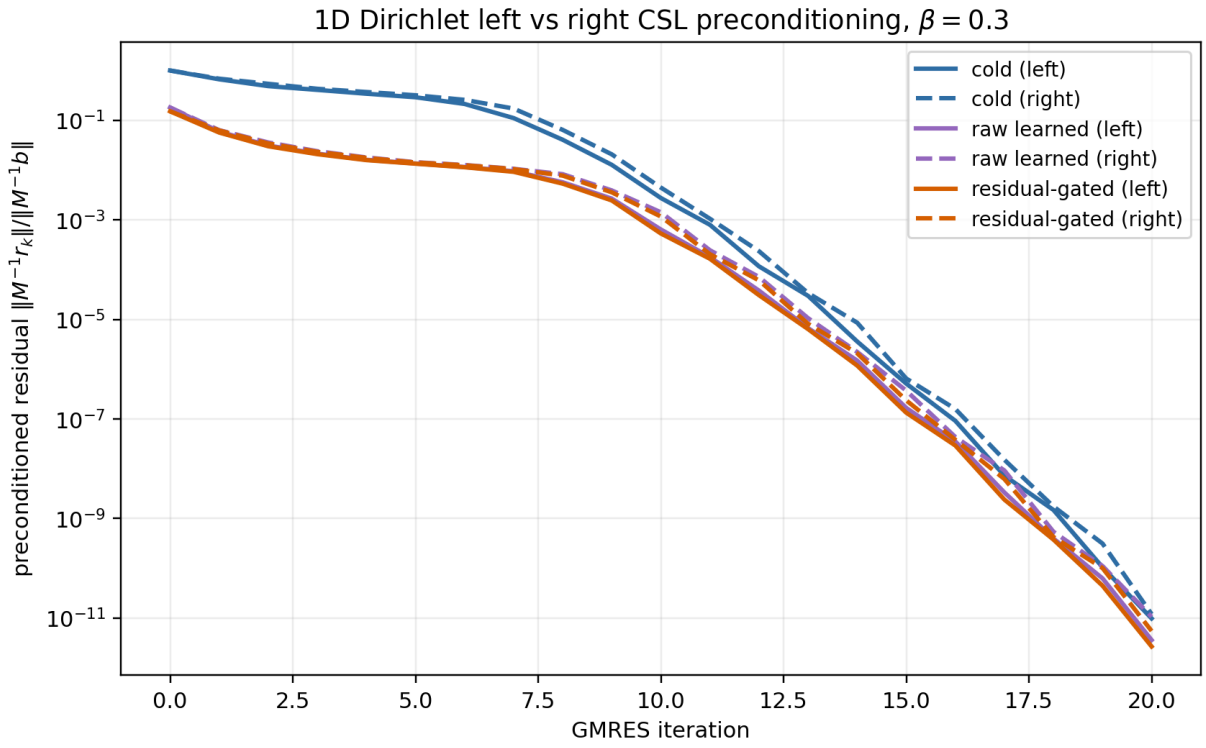


Figure 7.8: Preconditioned residual for the same comparison. Left preconditioning gives smaller preconditioned residuals because that is the quantity it directly minimises.

7.5. One-dimensional PML

The one-dimensional PML experiment bridges the clean Dirichlet analysis and the two-dimensional FD/PML problem. The operator is no longer symmetric and real. The PML stretching introduces complex-valued entries and removes the clean eigenbasis. This section therefore reports only residual and error diagnostics. No mode-by-mode analysis is possible.

The key question is whether the learned field remains operator-compatible when the absorbing layer is included. There are two distinct things to test. First, whether the network predicts the PML values accurately as a field. Second, whether the full FD/PML operator residual is small when the predicted field (PML included) is used as the initial guess. These can come apart. A prediction accurate in the physical interior but inconsistent in the absorbing layer will produce a large full-grid residual through the interface stencil, even if the interior looks acceptable.

Table 7.5: One-dimensional PML initial-state diagnostics, $16 \to 32$, $\beta = 0.3$. Full-grid FD/PML training simultaneously improves field error, true residual, and preconditioned residual. Zeroing the PML output of the analytic diagnostic model increases the true residual from 1.924 to 37.14, a $19\times$ amplification caused entirely by the interface mismatch.

Initial guess	Field error	$\|r_0\|_2/\|b\|_2$	$\|M_{\mathrm{CSL}}^{-1} r_0\|_2/\|M_{\mathrm{CSL}}^{-1} b\|_2$
Zero start	1.000	1.000	1.000
Analytic-PML raw	0.257	1.924	0.339
Analytic-PML interior-only	0.521	37.14	0.753
Full-grid FD/PML-trained	0.022	0.539	0.034

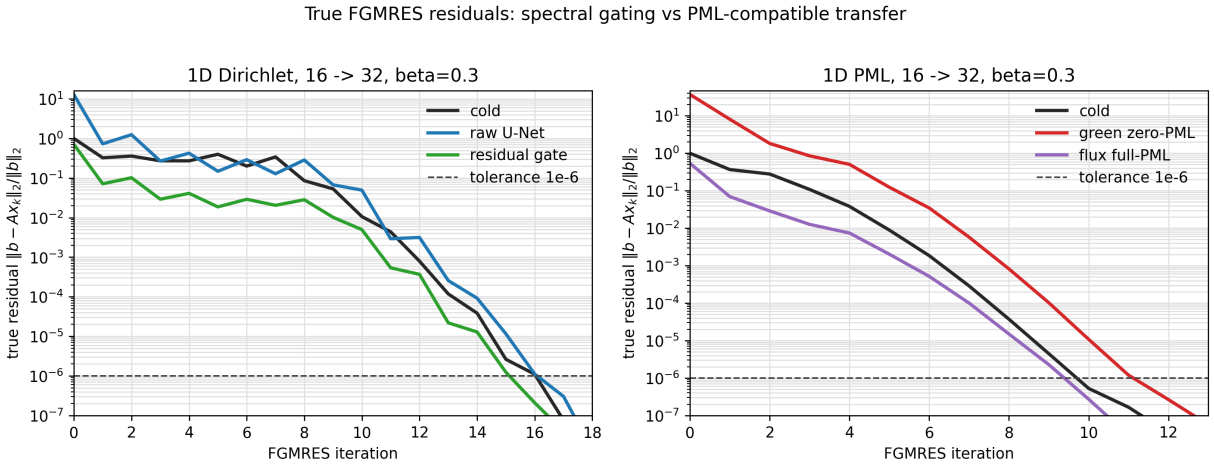


Figure 7.9: Side-by-side true-residual FGMRES curves for the two 1D experiments, $\beta = 0.3$. Left (Dirichlet): the raw learned start enters with a large true residual but the transient is short-lived; residual gating eliminates the spike. Right (PML): the full-grid FD/PML-trained model starts below the zero-start residual; zeroing the PML output produces a catastrophic residual spike, confirming that operator compatibility requires full-grid prediction.

Table 7.5 gives the key comparison. Zeroing the PML output increases the true residual from 1.924 to 37.14 for the analytic-PML model, a $19\times$ amplification caused entirely by the artificial interface discontinuity. This amplification is the spatial analogue of the residual-amplification identity in Proposition 5.1: the zeroed prediction introduces a discontinuity at the interior–PML interface whose image under A_{ω_H} falls in directions of large $|\lambda|$ and inflates the residual. The full-grid FD/PML-trained model avoids this by predicting the absorbing layer as part of its output and is simultaneously compatible with the full operator used to compute the residual. It achieves the lowest field error (0.022), the lowest true residual (0.539, below zero-start), and the lowest preconditioned residual (0.034). The conclusion is not that the PML must be

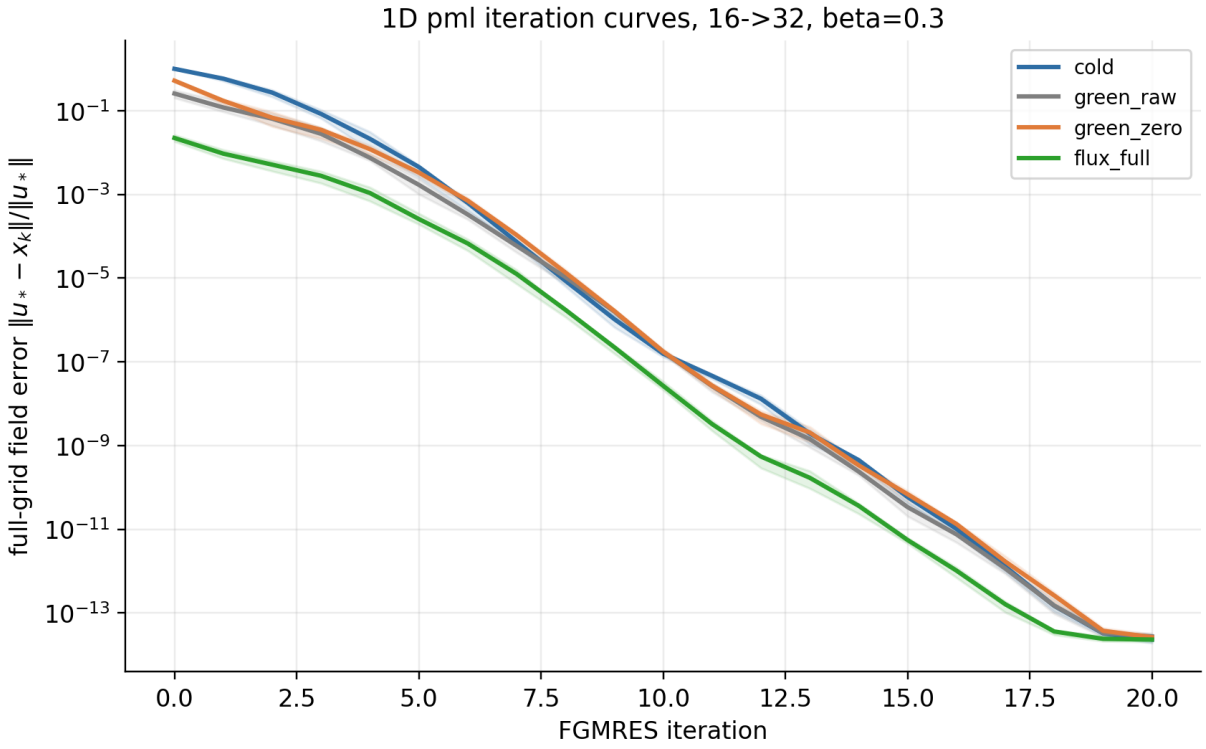


Figure 7.10: Field error versus FGMRES iteration, 1D PML experiment. The full-grid FD/PML-trained model starts close to the exact full-grid solution. The analytic-PML diagnostic starts are less accurate in field norm and substantially worse in residual norm.

visually accurate for its own sake, but that the network output must be compatible with the full operator used to compute the residual.

Together the Dirichlet and PML experiments identify two necessary conditions for solver-useful warm starts. A learned initial guess must be spectrally safe, in the sense that its error should not concentrate heavily in large-$|\lambda_k|$ components, and it must be operator-compatible on the full grid including the absorbing layer. These conditions are diagnostic rather than sufficient. Satisfying them improves the initial algebraic state, but does not by itself guarantee a large reduction in total FGMRES iterations.

7.6. Two-dimensional FD/PML results

The two-dimensional experiments test whether the one-dimensional criteria survive in the realistic setting. All results use the full-grid FD/PML-trained `TransferUNet` (`base_ch`=32, five levels) trained on 9600 samples per pair with CSL damping $\beta = 0.3$.

7.6.1. Fixed-budget evaluation

Ten independently generated test right-hand sides per pair are evaluated over 40 FGMRES iterations. No method reaches the convergence tolerance within this budget, so the comparison is on the final true and preconditioned residuals. The PML ratio,

$$\text{PMLRatio}(x_0) = \frac{\|x_0\|^2_{\text{PML}}}{\|x_0\|^2_{\text{interior}}}, \tag{7.6}$$

measures initial-guess energy in the absorbing layer relative to the physical interior. Interior-only starts have PML ratio 0 by construction.

Table 7.6 gives the central two-dimensional result. The PML ratio is not reported for the zero start because the identically zero initial guess has zero energy in both the PML and the physical interior, making the ratio undefined. The full-grid FD/PML-trained model used raw gives the lowest final true residual for

Table 7.6: Two-dimensional FD/PML fixed-budget evaluation, $\beta = 0.3$, 40 FGMRES iterations, 10 test samples per pair. All entries are sample means. The full-grid FD/PML-trained raw start gives the lowest final true residual for every pair. The preconditioned residual picture is more complex: at $64 \to 128$ the raw start produces a large preconditioned residual despite its best true-residual outcome, indicating incomplete compatibility with the CSL-preconditioned Krylov coordinates.

Pair	Method	Full error	PML ratio	Init. true res.	Init. prec. res.	Final true res.	Final prec. res.
$16 \to 32$	Zero start	1.000	–	1.00	1.00	2.05×10^{-3}	1.56×10^{-3}
	Interior-trained, interior-only	0.570	0.000	15.76	0.233	9.34×10^{-4}	4.32×10^{-4}
	Full-grid FD/PML raw	0.238	0.409	1.49	0.116	4.22×10^{-4}	2.30×10^{-4}
	Full-grid FD/PML interior-only	0.573	0.000	15.58	0.236	9.21×10^{-4}	4.62×10^{-4}
$32 \to 64$	Zero start	1.000	–	1.00	1.00	3.83×10^{-1}	3.08
	Interior-trained, interior-only	0.590	0.000	11.25	0.273	4.37×10^{-1}	0.764
	Full-grid FD/PML raw	0.352	0.365	1.93	0.712	3.28×10^{-1}	1.09
	Full-grid FD/PML interior-only	0.587	0.000	10.55	0.265	4.21×10^{-1}	0.749
$64 \to 128$	Zero start	1.000	–	1.00	1.00	4.33×10^{-1}	10.92
	Interior-trained, interior-only	0.616	0.000	6.40	0.895	5.22×10^{-1}	6.97
	Full-grid FD/PML raw	0.276	0.322	1.86	18.41	3.79×10^{-1}	28.51
	Full-grid FD/PML interior-only	0.542	0.000	5.93	0.685	4.66×10^{-1}	5.31

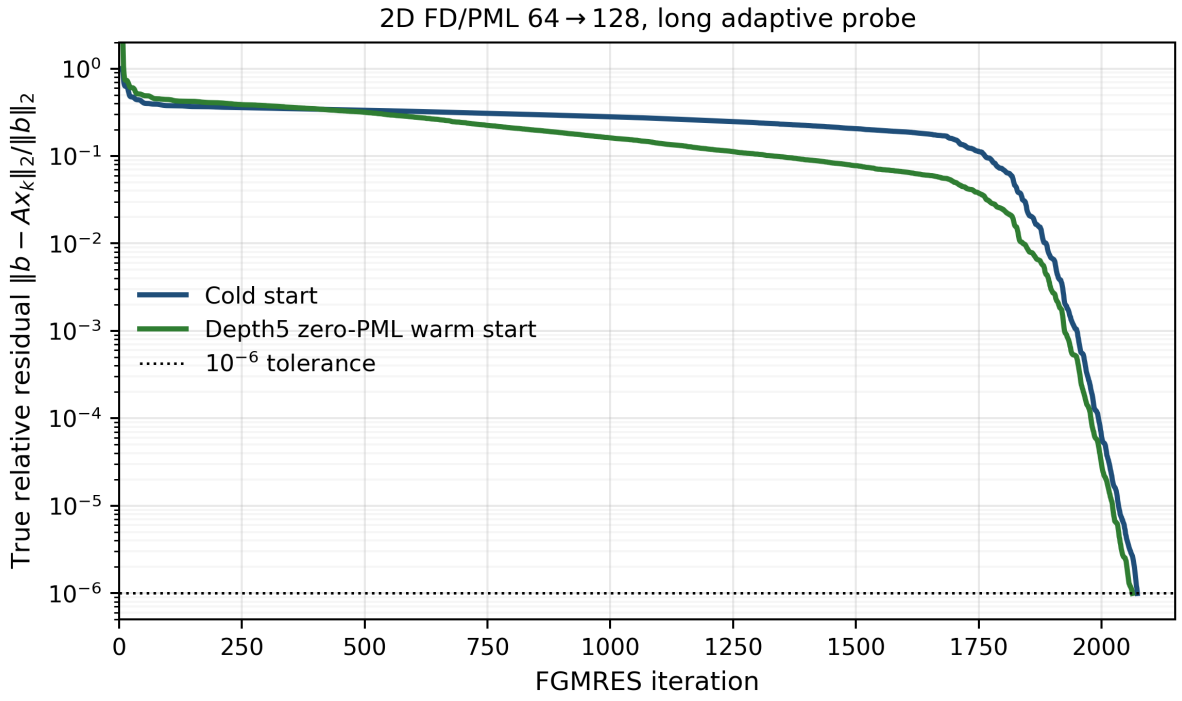


Figure 7.11: CSL–FGMRES convergence curves for the $64 \to 128$ pair, $\beta = 0.3$. The learned warm start achieves steeper early descent, reducing the finite-budget residual ($4.33 \to 3.79 \times 10^{-1}$) and delivering faster convergence for moderate stopping criteria.

all three pairs. The improvement is largest for $16 \to 32$ ($2.05 \to 0.422 \times 10^{-3}$, a $5\times$ reduction), and remains consistent at higher frequencies ($3.83 \to 3.28 \times 10^{-1}$ for $32 \to 64$; $4.33 \to 3.79 \times 10^{-1}$ for $64 \to 128$).

The role of full-grid training is visible in the raw versus interior-only comparison. For $16 \to 32$, the full-grid raw start has initial true residual 1.49, whereas the same model with its PML strip removed has initial true residual 15.58. This reproduces in 2D the interface-mismatch effect identified in the 1D PML experiment. The conclusion is not that every predicted PML value is physically meaningful, but that discarding the PML output makes the initial guess incompatible with the full FD/PML operator.

The preconditioned residual tells a more nuanced story, and the $64 \to 128$ row is where it concentrates. The raw start there achieves the best final true residual but produces a final preconditioned residual of 28.51, far above the zero-start value of 10.92. This is not an experimental artefact. An independent rerun using the frozen checkpoint, seed, ten test right-hand sides, and forty-iteration budget reproduced a final preconditioned residual of 28.5127, which rounds to the reported 28.51. The network minimises a field loss; the CSL-preconditioned Krylov process minimises a residual in coordinates the training never saw; and at high frequency the two coordinate systems pull apart.

This is the predicted consequence of the residual-amplification identity in Proposition 5.1: a field-trained model is not constrained to operator-compatible directions, so its remaining error concentrates in the modes the preconditioned system amplifies most strongly. The $64 \to 128$ result is therefore reported as a characterisation of where field-loss training is structurally insufficient, not as a failure of the warm-start concept. The residual-aware training direction that follows from this observation is discussed in Chapter 9.

Answering the research sub-questions. The two sub-questions of Chapter 2 can now be answered directly. SQ1 asked how accurately upward transfer predicts the high-frequency field and how that accuracy degrades across pairs: the full-grid field error rises from 0.240 at $16 \to 32$ to 0.339 at $64 \to 128$ (Table 7.2), a graceful degradation confirming the map is learnable at every tested pair. SQ2 asked whether field accuracy alone predicts solver acceleration: it does not. The same-run comparison of field error against the true and preconditioned residuals shows that a field-accurate start can be residual-incompatible, the $64 \to 128$ row being the clearest case, where the most field-accurate start produces the worst preconditioned residual. The preconditioned residual is the more faithful indicator of solver value. A systematic correlation of field error and initial residual against iteration savings across the full training sweep, and a direct comparison against an operator-aligned training objective, would quantify this further and are identified as future work in Chapter 9.

7.6.2. Residual versus field error: the 2D picture

The one-dimensional diagnostics rest on an explicit eigenbasis that the two-dimensional FD/PML operator does not provide, so the same field-versus-residual separation has to be shown here by direct measurement rather than by modal decomposition.

The one-dimensional eigenbasis analysis is not available in 2D. The FD/PML operator is complex-valued, non-normal, and does not admit the clean decomposition of Equation (6.5). Figure 7.12 therefore provides a qualitative analogue rather than a modal proof. The same separation between field error and residual behaviour is visible: learned starts begin closer to the exact solution in field norm, the PML interface mismatch inflates the true residual, and the Krylov correction is driven by residual reduction rather than monotone field-error decrease. The figure supports the transfer of the diagnostic lesson from 1D to 2D, but it does not provide a complete spectral explanation of the 2D FD/PML operator.

7.6.3. Adaptive convergence to tolerance

The fixed-budget experiment compares finite-budget residuals but does not answer how many iterations are needed to reach a prescribed tolerance. Adaptive runs were therefore conducted with stopping rule

$$\frac{\|b - Ax_k\|_2}{\|b\|_2} \leq 10^{-6}. \tag{7.7}$$

The $16 \to 32$ pair uses ten test right-hand sides; $32 \to 64$ uses three (owing to the much higher iteration count); and $64 \to 128$ is a single long-convergence probe, reported as a scale estimate rather than a sample mean. Learned starts are interior-only in these runs, isolating the value of the physical-domain prediction without the PML compatibility complications.

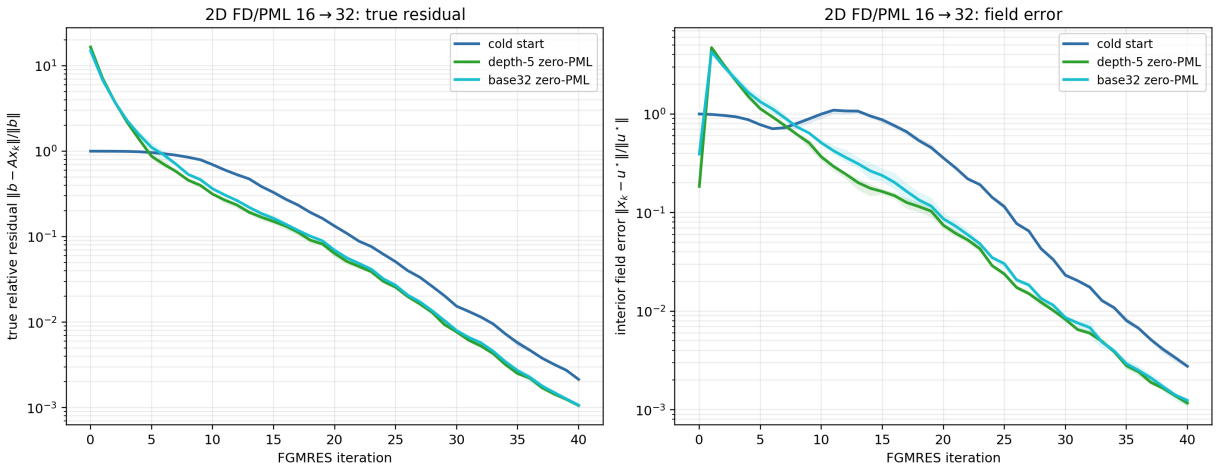


Figure 7.12: Same-run 2D FD/PML diagnostic, $16 \to 32$. Both panels use the same five test right-hand sides and colour scheme; shaded bands show the interquartile range. The learned starts begin closer to the exact solution in field norm but with large true residuals because zeroing the PML introduces a full-operator mismatch. The first FGMRES correction reduces the true residual and can simultaneously move the iterate farther from the exact solution in field norm. This is consistent with the 1D Dirichlet analysis but now visible without access to an eigenbasis.

The two learned interior-only rows compare two network widths of the same `TransferUNet` family. *Base32* is the `base_ch` $= 32$ model used throughout the chapter; *Base48* is the wider `base_ch` $= 48$ variant, included here to check whether additional capacity changes the convergence-to-tolerance saving. Both are evaluated interior-only.

Table 7.7: Adaptive 2D FD/PML convergence, $\beta = 0.3$, stopping tolerance 10^{-6}. Mean iteration count, with relative saving against zero start in parentheses. The $16 \to 32$ and $32 \to 64$ means are over ten and three right-hand sides respectively; the $64 \to 128$ column is a single long-probe estimate.

Start	$16 \to 32$	$32 \to 64$	$64 \to 128$
Zero start	69.4	347.3	2073
Interior-trained, interior-only	66.9 (3.6%)	343.0 (1.2%)	2062 (0.5%)
Base32 interior-only	67.0 (3.5%)	343.7 (1.0%)	2066 (0.3%)
Base48 interior-only	67.1 (3.3%)	343.0 (1.2%)	2070 (0.1%)

Table 7.7 and Figure 7.13 give the adaptive result. All learned starts reach the tolerance. The savings are positive but small: 3.6% for $16 \to 32$, 1.2% for $32 \to 64$, and 0.5% in the single $64 \to 128$ probe. The relative saving shrinks with frequency because the absolute iteration count grows sharply, from 69 to 2073, while the learned start contributes a fixed initial-state correction that does not scale with the problem difficulty.

The fixed-budget and adaptive results are complementary. The fixed-budget table tests which start gives the best residual after a fixed number of iterations, while the adaptive experiment tests how many iterations are required to reach a prescribed tolerance. The full-grid raw start is best on the fixed-budget metric; conservative interior-only starts give small convergence-to-tolerance savings. Neither result claims a changed convergence mechanism. The learned start saves a small number of iterations by improving the initial state, and the asymptotic difficulty of the high-frequency CSL-preconditioned system is unchanged.

The small to-tolerance saving is not the only solver-relevant effect, and reading it as such understates the warm start. Over the fixed budget the descent is appreciably steeper. At $16 \to 32$ the full-grid warm start reaches a final true residual of 4.22×10^{-4} after forty iterations against 2.05×10^{-3} for the zero start, so over the same iterations it contracts the residual by roughly 3.5×10^{3} while the zero start contracts

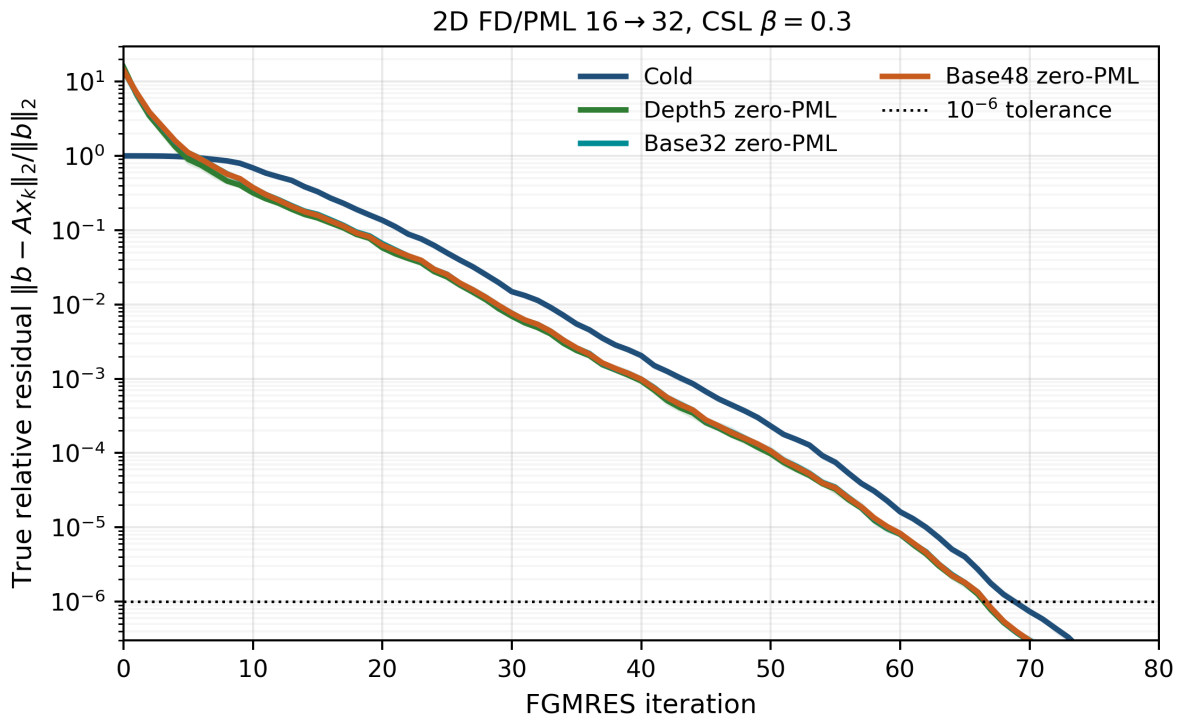


Figure 7.13: Mean true-residual convergence curves, adaptive $16 \to 32$ experiment. Sample mean over ten test right-hand sides with shaded interquartile bands. The learned interior-only starts reach tolerance a few iterations before the zero start on average, saving 3.5 to 3.6% of the total iteration budget.

it by roughly 4.9×10^2, even though it enters from a higher initial residual. The two trajectories share the asymptotic CSL-controlled slope, but the warm start spends fewer iterations in the early, slow phase before that slope is reached. A plausible explanation is that the warm start has already removed part of the error in the well-conditioned, large-$|\lambda_k|$ subspace that the CSL-preconditioned Krylov polynomial resolves first, so the iteration reaches its rapid phase sooner; what it cannot accelerate is the near-zero part of the spectrum that governs the tail, which is why the count to 10^{-6} barely moves.

This distinction is useful rather than cosmetic. Many settings in which a Helmholtz solve is only an inner step, such as the linearised solves inside a Gauss–Newton iteration of full-waveform inversion, do not require machine precision: a moderate relative tolerance of a few percent is often enough and is re-tightened only in later outer iterations. In that regime the relevant quantity is the iteration count to a loose tolerance, where the steeper early descent of the warm start gives a larger proportional saving than the to-10^{-6} figures suggest. The learned warm start is therefore most valuable exactly where it is most often deployed, as an accelerator of the early-to-moderate-accuracy phase rather than of the final tail.

The stopping rule of Equation (7.7) measures the true residual. Since the left-preconditioned iteration internally minimises the preconditioned residual, the iteration counts in Table 7.7 need not coincide with the savings obtained under a preconditioned-residual stopping rule, because the two norms can order intermediate iterates differently. A direct comparison is deferred to future work.

7.7. From one dimension to two and beyond

The progression from Dirichlet to PML to two-dimensional FD/PML is a deliberate ladder, and the same logic motivates an eventual extension to three dimensions.

Two conclusions transfer directly across dimensions. The identity $r_0 = A_{\omega_H} e_0$ is dimension-independent, so the distinction between field error and operator residual persists in 3D. The role of the warm start as an initial-state correction that leaves the CSL-preconditioned operator unchanged is structural and is not specific to the dimension.

Several effects become harder in 3D. The absorbing layer surrounds the domain on more faces, increasing the relative weight of interface-compatibility errors in the full-grid residual. For the fixed-grid configuration of this thesis ($n = 512$, $n_{\text{pml}} = 112$ per side), the physical interior spans $(288/512)^2 \approx 32\%$ of the total 2D grid area; the analogous 3D configuration at the same resolution and PML width would reduce this to $(288/512)^3 \approx 18\%$, so operator-compatibility of the absorbing layer carries proportionally larger weight in any full-grid residual evaluation. Source interference becomes more intricate as outgoing wavefronts overlap over a larger volume, enlarging the globally coherent error component identified in the source-wise diagnostics. Memory cost grows with the number of grid points, and the fixed-n configuration used here does not obviously scale to a three-dimensional grid at matched accuracy. Krylov iteration growth is expected to be at least as severe in 3D as in 2D (CSL-preconditioned convergence degrades at least linearly with frequency in two dimensions [11], and the indefinite spectral band that limits convergence widens further with dimension) so the absolute iteration budget against which any warm-start saving is measured will grow disproportionately relative to the fixed initial-state correction the learned operator provides.

The two-dimensional experiments are therefore a necessary intermediate step. They retain the analytical transparency of the 1D setting, including direct residual diagnostics, explicit PML comparison, and interpretable field structures, while exposing the FD/PML compatibility and interference effects that any three-dimensional study must confront. This dimensional discussion should be read as a scaling argument, not as a demonstrated 3D result. The experiments in this thesis stop at 2D FD/PML. The 3D comments identify which mechanisms are expected to persist and which costs are expected to grow.

7.8. Chapter summary

Five conclusions follow from the experiments.

Field accuracy is not solver accuracy. The modal identity $c_k(r_0) = \lambda_k c_k(e_0)$ explains why: field error in large-$|\lambda_k|$ modes is amplified into residual. A learned start with RelL2 $= 0.084$ can still produce a true residual of $11.6\|b\|_2$. The resulting transient is short-lived inside CSL-FGMRES, but it confirms that field-prediction metrics alone cannot certify solver usefulness. The preconditioned residual, which is closer to the field error than the true residual is, gives a more faithful read of the learned start's solver value.

Operator compatibility requires full-grid prediction. Zeroing the PML output amplifies the true residual by up to $19\times$ in 1D and produces initial true residuals of 15 to $16 \times \|b\|_2$ in 2D. The absorbing layer is part of the discrete operator, not a passive boundary decoration. Among the tested variants, full-grid FD/PML training is the only approach that consistently improves all three initial-state metrics (field error, true residual, and preconditioned residual) relative to the zero start.

The network learns a source-wise decomposition without supervision. In the analytic multi-source diagnostic, field-only training produces a CNN whose predictions align with source-centred outgoing contributions, even though source labels are not provided. Its errors concentrate in the globally coherent interference terms that local source-wise models cannot represent, quantified by a $1.9\times$ error ratio between high and low omitted-interference regions. The amplitude-ablation experiment rules out the passive Green's-function interpretation: removing the $1/\sqrt{r}$ amplitude singularity collapses prediction quality to the zero predictor, so the source-wise structure is a learned feature, not a kernel side-effect. This is the primary scientific finding of the chapter; it is offered as a mechanism explanation for the solver behaviour, not as a direct claim about the FD/PML model.

Warm-start benefits are positive but modest. In the 2D fixed-budget experiments the full-grid warm start reduces the finite-budget true residual for all three pairs, with the largest improvement at $16 \to 32$ ($5\times$ final true-residual reduction). Adaptive convergence-to-tolerance runs give iteration savings of 3.6%, 1.2%, and 0.5% for the three pairs respectively. The savings shrink with frequency because the absolute iteration count grows while the initial-state contribution remains fixed.

The learned model is a warm start, not a preconditioner. The $64 \to 128$ fixed-budget result formalises this distinction. The raw warm start achieves the best final true residual while producing the worst final preconditioned residual, the predicted consequence of Proposition 5.1: the field-trained model is not constrained to operator-compatible directions, so its remaining error concentrates in the modes the CSL-preconditioned coordinates amplify most strongly. CSL handles the large eigenvalues. The near-zero part of the spectrum, which the learned operator targets, is precisely the regime where CSL alone

is weakest. Closing the remaining gap requires training against a residual-aware objective rather than a field-only one. Chapter 9 discusses this direction.

8

Discussion

The headline result of Chapter 7 can be stated in one sentence: in the present experiments the absolute number of FGMRES iterations does not collapse, but the large eigenvalues of the indefinite Helmholtz operator are already handled by CSL, and the near-zero part of the spectrum, which is precisely where CSL alone is weakest, is the part the learned operator influences. This is not a factor-of-ten solver speedup, and the thesis does not claim one. It is the kind of result that identifies a useful ingredient and the place in the algorithm where that ingredient belongs. The finite-budget true-residual improvement at all three frequency pairs, and the emergence of a source-wise decomposition that the network was never asked to learn, are the two pieces of evidence that the ingredient is real.

This chapter interprets that evidence. Section 8.1 gives the algebraic reason field accuracy and solver usefulness are not equivalent. Section 8.2 interprets the emergent source-wise structure as the primary scientific contribution and discusses what it implies for how learned wave methods should be designed. Sections 8.3 and 8.4 separate the claims the evidence supports from the claims it does not. Section 8.5 positions the work against classical alternatives that exploit the same linear structure. Section 8.6 addresses training cost and amortisation. Section 8.7 records the main threats to validity. Section 8.8 gives the closing interpretation.

8.1. The residual-amplification mechanism

The central object is the operator-amplified residual

$$r_0 = b - A_{\omega_H} x_0 = A_{\omega_H} e_0, \qquad e_0 = u_\star - x_0, \tag{8.1}$$

where $u_\star$ is the exact discrete high-frequency solution and x_0 is the learned initial guess. The field error e_0 is the quantity minimised by the network. The residual r_0 is the quantity seen by FGMRES. These two objects are linked by A_{ω_H}, and this link is the source of the observed turn-on transient.

In the one-dimensional Dirichlet eigenbasis, the relation becomes the exact modal identity

$$c_k(r_0) = \lambda_k \, c_k(e_0), \tag{8.2}$$

as derived in Section 5.5 and measured directly in Table 7.3. Classical spectral analyses of the CSL-preconditioned Helmholtz operator characterise where the eigenvalues sit and how tightly they cluster [28]. Equation (8.2) adds the complementary map: it tracks how a given field error is redistributed into the residual by those same eigenvalues, and therefore converts the spectral picture into an explicit error-to-residual budget.

A learned start with field error 0.084 and true residual $11.64\|b\|_2$ is not a numerical failure. It is the predicted behaviour of a model whose remaining error has components in modes with large $|\lambda_k|$. The solver does not ask whether the field looks close. It asks whether applying A_{ω_H} to the error produces a small residual.

This mechanism gives a concrete modal safety condition. A learned warm start improves on the zero-start residual in mode k only if

$$|\lambda_k| \, |c_k(e_0)| < |c_k(b)|. \tag{8.3}$$

The right-hand side is the zero-start residual coefficient. The left-hand side is the learned field error after amplification by the high-frequency operator. Because $|\lambda_k|$ varies strongly across the spectrum, the permitted field error is highly non-uniform across modes. A standard solution-norm loss does not encode this non-uniform budget. It can therefore reduce field error while leaving too much error in residual-sensitive directions.

Equation (8.3) is the main analytical interpretation of the results. It is architecture-independent. The error distribution may come from a CNN, a U-Net, a Fourier neural operator, a classical interpolant, or a hand-built continuation rule, but the operator decides how that error is converted into a residual. The mechanism also makes the claim falsifiable: a future learned model that violates this modal budget cannot be expected to be a reliable warm start, even if its field error is small.

The PML experiments show the spatial analogue of the same principle. In the Dirichlet setting, incompatibility is visible through modal amplification. In the PML setting, it is visible through the full-grid stencil and the absorbing layer. Removing or mispredicting the PML output changes the residual on the complete computational grid, even if the physical interior looks reasonable. The 1D PML diagnostic in Table 7.5 makes this concrete: zeroing the PML output of the analytic model amplifies the true residual by $19\times$, while a full-grid FD/PML-trained model achieves true residual 0.539 and preconditioned residual 0.034. The common conclusion across both settings is that the solver-facing object is not the plotted field, but the residual produced by applying the high-frequency discrete operator to that field.

8.2. The emergent source-wise decomposition

One of the most striking findings of the thesis is not in the solver curves. In Figures 7.2 and 7.3, a convolutional network trained purely against a field loss, with no source labels, source counts, or structural supervision, produces source-centred outgoing wave packets whose error structure is consistent with a Voronoi-windowed interpretation of the source configuration. The error increases monotonically with the omitted-interference field (Figure 7.4), by a factor of 1.89 between the lowest and highest quintiles. The amplitude ablation causes prediction quality to collapse towards the zero-predictor baseline, showing that amplitude information is necessary in this diagnostic and arguing against a purely phase-driven interpretation.

This result deserves interpretation on its own terms. Three observations follow.

First, this is sparse dictionary learning in a setting where no dictionary was provided. The network is in effect discovering that the solution at the target frequency admits a compressed representation as a sum of source-centred outgoing contributions, each localised to one cell of the Voronoi partition of the source configuration. Classical sparse-coding methods such as K-SVD assume a fixed set of localised basis functions and infer their coefficients from data [35]. Here no such structure is assumed or inferred explicitly: it is implicit in the trained weights, but recoverable through the Voronoi diagnostic. The interpretability foil constructed in Section 7.3 is what makes that structure visible. The same Green's-function structure has recently been built into neural operators by construction, where an explicit finite-dimensional Green's representation is learned and then used to precondition iterative solvers [17]. The finding is the complementary observation: a convolutional network recovers an equivalent source-wise structure without being given the Green's formulation, purely from a field loss. This diagnostic complements standard neural-PDE interpretability tools. Rather than attributing a prediction through input-gradient saliency or assuming an explicitly learned basis, it tests whether the observed error structure is consistent with a local partition-of-unity representation. The amplitude ablation is a controlled diagnostic of the information needed for that representation; it does not by itself prove that the network explicitly constructs Voronoi cells. The connection between the approximation-theoretic and spectral-amplification perspectives can be stated precisely: the Voronoi-windowed approximation of Equation (7.3) is a partition-of-unity decomposition into source-centred outgoing contributions, and by Proposition 5.1 the omitted-interference term $\delta(x)$ (the component no local source-wise term can represent) is globally coherent and spatially extended, projecting preferentially onto Dirichlet modes with large $|\lambda_k|$. A field-only training objective is therefore permissive in exactly the modal subspace the high-frequency operator amplifies most strongly into residual: the frame representation error and the residual-amplification zone coincide, and partition-of-unity completeness at the field level does not imply residual safety at the solver level.

Second, the result connects directly to the residual-amplification mechanism of Section 8.1. The omitted-interference field is, by construction, the part of the high-frequency solution that no purely lo-

cal source-wise model can represent. It is globally coherent and spatially extended. Such fields tend to project onto modes with large $|\lambda_k|$. The network therefore concentrates its residual error precisely in the directions that the operator amplifies most strongly. This is not a coincidence: the field loss is permissive in exactly the modal directions the operator is sensitive to, and the Voronoi structure is the geometric image of that permissiveness. It is the same mechanism viewed from a different angle.

Third, the finding suggests that the next architectural step for learned wave methods is not larger models or different backbones. It is matching the loss to the operator. A field-only loss permits residual-incompatible solutions. A residual-aware loss, or a loss that weights modal components according to their eigenvalue magnitude, would penalise the precise error distribution the Voronoi diagnostic reveals. The architecture itself is already learning something physically interpretable. The training objective is what is permitting the operator-incompatible residual to remain. The same conclusion has been reached independently for hybrid neural-classical solvers, where reliability was found to depend on whether the training objective is aligned with the solver dynamics rather than on the neural architecture [36].

This finding carries one important qualification. The Voronoi diagnostic uses an analytic multi-source model trained on free-space Green's function targets, not the FD/PML model used in the solver experiments. The two-dimensional solver-facing model is evaluated through residuals and FGMRES curves, not through Voronoi overlays. The relevance of the finding to the FD/PML model is therefore mechanistic rather than direct: it identifies an error structure that is consistent with the residual behaviour observed in the solver experiments, but it does not claim that the FD/PML model performs an explicit source decomposition. Establishing that would require constructing the analogous diagnostic for the FD/PML model, which is beyond the present scope.

8.3. What the method achieves

Three positive claims survive critical inspection.

First, nontrivial cross-frequency structure is learnable. The two-dimensional full-grid FD/PML-trained TransferUNet reaches full-grid validation relative ℓ^2 errors of approximately 0.240, 0.321, and 0.339 for $16 \to 32$, $32 \to 64$, and $64 \to 128$ respectively. These errors are well below the zero predictor. The network has therefore learned meaningful frequency-transfer structure. This is a learning success, but not yet a solver success, because the loss is measured in field norm while FGMRES is governed by residuals.

Second, finite-budget true residuals improve in the two-dimensional FD/PML experiments. No method reaches the prescribed tolerance within the forty-iteration budget at $\beta = 0.3$, so the fixed-budget residual is the most informative comparison. The full-grid FD/PML-trained warm start gives the lowest final true residual for all three frequency pairs. The headline changes are summarised in Table 8.1.

Table 8.1: Fixed-budget true-residual improvement of the full-grid FD/PML-trained warm start in the two-dimensional FD/PML experiments. Values are the sample means reported in Chapter 7 after forty FGMRES iterations.

Pair	Zero start	Full-grid learned start
$16 \to 32$	2.05×10^{-3}	4.22×10^{-4}
$32 \to 64$	3.83×10^{-1}	3.28×10^{-1}
$64 \to 128$	4.33×10^{-1}	3.79×10^{-1}

The improvement is largest at the lowest frequency pair ($\approx 5\times$ reduction) and consistent but smaller at the higher pairs. The method is therefore useful as a warm start in the tested setting, where useful means finite-budget residual improvement. This is weaker than convergence acceleration to tolerance, but it is stronger than mere visual field agreement.

Third, full-grid FD/PML training removes the absorbing-layer incompatibility of interior-only models. An interior-only model can produce harmful output in the absorbing layer, and zeroing the PML strip then acts as a safety patch. The full-grid FD/PML-trained model is trained against the full FD/PML solution including the absorbing layer, and keeping its PML output gives better final true residuals than zeroing

it. The methodological lesson is that the PML must be learned as part of the operator data. It is not an irrelevant boundary decoration.

The Voronoi finding of Section 8.2 adds a fourth, qualitatively different positive claim: the network learns a physically interpretable structure that was not part of the training signal. This is the scientific contribution that is most likely to outlast the specific solver numbers reported here.

8.4. What the method does not achieve

The same evidence imposes sharp limits.

A warm start changes r_0, but it does not change the spectrum or pseudospectrum of the preconditioned operator. Once the initial residual has been processed, the remaining convergence is governed by the classical CSL-preconditioned FGMRES problem [25]. This explains why early or finite-budget residual behaviour can improve without a large reduction in the iteration count. The present method should therefore not be described as a factor-of-ten solver acceleration.

The current learned correction is a warm start, not a preconditioner. The strongest direct evidence is the $64 \rightarrow 128$ row of Table 7.6: the raw full-grid learned start achieves the best final true residual (3.79×10^{-1}) while simultaneously producing the worst final preconditioned residual (28.51, versus the zero-start value of 10.92). This is not a contradiction. The network minimises a field loss; the CSL-preconditioned Krylov process minimises a residual in coordinates the training never saw; and at high frequency the two coordinate systems pull apart. This is exactly what the residual-amplification identity of Proposition 5.1 predicts: a field-trained model is not constrained to operator-compatible directions, so its remaining error concentrates in the modes the preconditioned system amplifies most strongly. The output is not uniformly compatible with the residual seen by the CSL-preconditioned Krylov process. This is why the term warm start is retained. The current model is not a learned preconditioner. Constructing such a learned preconditioner, in which a network maps residuals to corrections inside the Krylov iteration, is precisely the direction realised for the heterogeneous Helmholtz equation by [32, 33]; the present warm-start study isolates the residual-compatibility condition that any such preconditioner must satisfy at every step.

The convergence-to-tolerance savings are small. Adaptive runs at tolerance 10^{-6} give iteration savings of 3.6% ($16 \rightarrow 32$), 1.2% ($32 \rightarrow 64$), and 0.5% ($64 \rightarrow 128$, single-probe). The savings shrink with frequency because the absolute iteration count grows sharply (from 69 to 2073) while the learned start contributes a fixed initial-state correction that does not scale with the problem difficulty. A method that saves a constant number of iterations cannot, by itself, deliver order-of-magnitude solver acceleration in the high-frequency limit.

High frequency is the stress test of these limitations. As ω increases, the spectrum of A_{ω_H} becomes more indefinite and the band of large $|\lambda_k|$ widens. The modal field-error budget in Equation (8.3) tightens. This explains why the best relative improvement occurs for the lowest tested pair and weakens at higher frequencies.

The thesis therefore does not claim that learned frequency transfer replaces CSL, nor that it already defines a production-ready learned preconditioner. The claim is narrower: learned frequency transfer can produce useful warm-start information in controlled FD/PML settings, but only when the prediction is compatible with the residual seen by the high-frequency operator.

8.5. Comparison with classical alternatives

The strongest criticism of learned frequency transfer in the linear homogeneous setting is not that the network fails to learn. It is that classical methods already exploit linearity, superposition, and Green's-function structure very efficiently. A fair comparison must therefore be against classical methods that use the same structure, not against weak unpreconditioned Krylov baselines.

CSL is complementary rather than competitive. It transforms the indefinite problem into a shifted system, and the learned model only changes the initial guess supplied to the CSL-preconditioned Krylov method [9]. Any claim that the network beats CSL would be inaccurate, because CSL remains the preconditioner in the successful experiments.

For a fixed homogeneous linear operator, the strongest baselines exploit the solution structure directly. A Green's-function quadrature or spectral expansion can approximate the solution from a modest number

of precomputed responses, preserves linearity exactly, and may require far fewer than the 9600 supervised solves used per frequency pair in this thesis.

When an explicit Green's function is unavailable, fixed wave operators often have low-rank off-diagonal blocks that randomised probing can exploit with a number of solves proportional to the numerical rank [31]. Domain-decomposition preconditioners that propagate polarised traces between layers reach near-linear complexity on heterogeneous two-dimensional models without any learned component [37]. Classical Krylov methods can also reuse information across nearby shifted systems through subspace recycling, rational approximation, and deflation, the last of which removes the near-zero eigenvalues that limit CSL by projecting them out explicitly [38, 39]. A neural model trained from scratch can therefore relearn structure that classical methods obtain almost for free, and the learned warm start is best read as a data-driven counterpart to these deflation and recycling strategies rather than as a competitor to them.

The learned approach has a plausible opening only where these classical shortcuts weaken. Relevant regimes include non-smooth sources distributed throughout the domain, near-source quantities where far-field moments are insufficient, many interacting parameters, tolerance targets closer to a few percent than to machine precision, and nonlinear wave equations without a Green's-function superposition principle. The Voronoi finding of Section 8.2 also points in this direction: the network's spontaneous source-wise decomposition is the kind of structural recovery that classical methods would have to be told about, whereas the network discovers it from data.

The linear FD/PML study is consequently not the application where learned transfer is expected to be most competitive. It is the baseline difficulty test. If a learned method cannot be made residual-compatible in the linear problem, it is unlikely to be reliable in nonlinear or high-parameter settings. The linear study is therefore best understood as a controlled diagnostic, not as a final deployment claim.

8.6. Training cost and amortisation

Training cost is a first-order limitation. The two-dimensional datasets use 9600 exact FD/PML solves per frequency pair. Such a cost is justified only when the trained model is reused across many later solves, or when the learned component enables a regime that classical methods cannot handle. The most favourable linear case is a fixed operator with many right-hand sides, where the medium, grid, boundary treatment, and frequency pair remain stable enough for reuse. If the medium changes after each solve, as in many full-waveform inversion workflows, the amortised value of a fixed trained model weakens.

A useful reporting standard follows from this observation. Solver acceleration cannot be judged from per-solve iteration count alone, because the total cost includes data generation, training, inference, and any classical preconditioner still used at deployment. A learned solver component should report the number of training solves, the cost of high-fidelity labels, the training cost, and the break-even number of deployment solves. The break-even condition can be written schematically as

$$N_{\text{break even}} = \frac{C_{\text{data}} + C_{\text{train}}}{C_{\text{zero}} - C_{\text{learned}}}, \tag{8.4}$$

where C_{zero} is the deployment cost of the classical zero-start solve and C_{learned} is the deployment cost when the learned component is used. This thesis does not claim that this break-even point is reached. It identifies the condition under which such a claim would have to be evaluated. At the lowest tested pair ($16 \to 32$), the 3.6% saving corresponds to approximately $\lfloor 0.036 \times 69 \rfloor \approx 2$ to 3 iterations saved per deployment right-hand side. With 9600 data-generation solves per frequency pair, Equation (8.4) then requires roughly $9600/2.5 \approx 3840$ deployment right-hand sides to recover data-generation cost alone, before accounting for training overhead or inference time. This is a conservative bound: at $64 \to 128$ the absolute iteration count is 2073, so the 0.5% saving corresponds to approximately 10 iterations saved per right-hand side, reducing the break-even requirement to roughly 960 deployment solves, a more favourable condition, though the single long-probe estimate does not yet verify it statistically. A seismic inversion workflow with $S \sim 300$ sources swept across several frequency pairs and repeated over multiple nonlinear iterations can plausibly exceed either threshold, making the present setting a candidate for amortised deployment in the regime where the medium and grid are fixed and right-hand sides are numerous.

The most compelling long-term cost argument is the nonlinear case. For nonlinear frequency-domain wave equations there is no Green's function and no superposition, so the solution cannot be assembled

from a small set of precomputed responses. Newton-type and Gauss–Newton methods are then standard, and their convergence depends on whether the initial state lies inside the basin of attraction of the desired solution. In that regime an initial guess is not merely an iteration-saving device. A poor initial state can prevent convergence entirely.

Frequency continuation is the classical multiscale remedy in full-waveform inversion: a converged low-frequency state initialises the next higher frequency [20]. A learned cross-frequency map can be viewed as a data-driven version of this continuation idea. A closely related strategy already uses neural networks to extrapolate the missing low frequencies that frequency continuation requires, which is the mirror image of the upward transfer studied here [16]. It may therefore become a feasibility mechanism rather than only a speed-up mechanism.

This argument must also be limited. Each Newton or Gauss–Newton step still requires the solution of a linearised wave problem. The residual-amplification constraint identified in this thesis therefore does not disappear. A learned initial state must be operator-compatible at each linearisation, not merely field-accurate at the outer nonlinear level. The open question is whether the classical-versus-learned difficulty gap narrows as nonlinearity and parameter dimension increase.

8.7. Threats to validity

Several factors bound the claims.

The physics is linear and homogeneous. This makes the analysis clean and the datasets reproducible, but it also gives classical methods the maximum possible advantage. The thesis therefore demonstrates a mechanism in a controlled baseline rather than competitiveness with Green's-function, quadrature, or low-rank methods.

The two-dimensional model is a warm start, not a preconditioner. It sets x_0 and then leaves CSL-FGMRES to solve the system. It is not applied as M^{-1} inside each Krylov iteration. A preconditioner claim would require residual-to-correction training on Krylov states and evaluation inside FGMRES.

The sample size and frequency coverage are limited. The main $\beta = 0.3$ solver evaluation uses ten test right-hand sides per pair and three dyadic pairs. This supports the qualitative ordering reported in Chapter 7, but not tight confidence intervals over source distributions, random seeds, or alternative frequency ratios. The interaction between the learned warm start and the CSL damping parameter is also not systematically explored.

The PML operator is non-normal. The Dirichlet modal analysis relies on an orthonormal eigenbasis, whereas the FD/PML operator is complex-valued and non-normal. Eigenvalues alone do not control transient behaviour or residual amplification. Pseudospectral effects can make a residual component large even when eigenvalue plots look benign [8]. This is why the PML and FD/PML experiments rely on direct residual diagnostics rather than on a modal interpretation alone.

The Voronoi diagnostic is mechanistic, not direct. It is obtained with a separate analytic multi-source model, and its quantitative numbers (monotone error rise, $1.89\times$ ratio between extreme quintiles) refer to that model. The relevance to the FD/PML solver-facing model is by mechanism, not by direct repetition of the diagnostic in the 2D setting. A future check would construct an analogous decomposition diagnostic for the FD/PML model.

The architecture coverage and discretisation are fixed. No two-dimensional solver-facing comparison against several neural-operator architectures is provided. Convolutional architectures are tied to a fixed grid and do not automatically define a convergent continuous operator under refinement, whereas neural-operator constructions are designed to remain meaningful across discretisations [40]. The residual-amplification threshold in this thesis is also defined at the training discretisation.

The research framing was initially tool-first. The project began by asking whether convolutional networks could learn frequency transfer. That framing risks applying a tool where classical structure is already superior. The value of the work is that this tool-first investigation exposed a problem-first mechanism: residual amplification is the constraint that any learned or classical transfer approximation must satisfy before it becomes solver-useful.

8.8. Closing interpretation

The thesis began from the observation that adjacent Helmholtz frequencies are not independent. A network can indeed learn part of the map between them. That observation is true, but incomplete. A Krylov solver does not consume visual similarity or field similarity. It consumes residuals. The important question is therefore not only whether the high-frequency field can be predicted, but whether the remaining prediction error lies in directions that the discrete operator and the CSL-preconditioned Krylov process can tolerate.

Equation (8.3) makes this tolerance condition explicit in the Dirichlet setting. The empirical answer is mixed in an informative way. Learned transfer contains solver-relevant information, but that information is not automatically residual-compatible when inserted as a raw warm start. The turn-on transient observed for raw learned starts is the evidence for this division. It shows that useful field information is present, but that it must be inserted into the solver in an operator-compatible form.

The Voronoi result then sharpens the interpretation. The network does not merely happen to produce field-accurate predictions. It organises those predictions into a source-wise decomposition that no part of the training signal asked for. The remaining error of that decomposition concentrates in globally coherent interference terms, which is also where the operator is most sensitive. The architecture is therefore producing exactly the structure the high-frequency operator finds hardest to tolerate. This is informative both as a scientific finding about what convolutional networks learn from wave data, and as a prescription for the next step: a training objective that respects the residual.

The final interpretation is therefore complementary rather than competitive. CSL remains the classical mechanism that stabilises the high-frequency Helmholtz solve. It controls the large eigenvalues. The learned operator contributes a different ingredient: cross-frequency information about wavefield geometry, including the source-wise structure that becomes visible only through the Voronoi diagnostic. As a one-time warm start, this ingredient improves finite-budget residual trajectories across all three tested frequency pairs, with the largest gain at the lowest frequency. It does not yet change the asymptotic convergence behaviour of the CSL-preconditioned Krylov process, and it does not yet act as a preconditioner. What it identifies, however, is the place in the algorithm where learning is the right tool: not at the large eigenvalues that CSL already handles, but at the near-zero part of the spectrum where classical preconditioners are weakest. That the near-zero eigenvalues are the binding constraint on CSL-preconditioned convergence is established for the classical solver by two-level deflation, which restores wavenumber-independent convergence precisely by projecting those modes out [41]. The learned operator is aimed at the same spectral region by different means.

The remaining gap between this signal and a deployed solver acceleration is the subject of Chapter 9.

9

Conclusion

This chapter records, in priority order, what the thesis establishes and what should be done next.

9.1. Conclusions

The central question of Chapter 2 was whether learned frequency transfer can produce solver-useful warm starts for CSL-preconditioned FGMRES at high frequency. The answer is a qualified yes: the warm start lowers the finite-budget true residual at all three pairs and gives small positive convergence-to-tolerance savings, but only when the prediction is operator-compatible, that is trained and evaluated on the full FD/PML grid the solver applies. The same condition governs any cross-frequency continuation.

C1. The network learns a source-wise decomposition without supervision. Trained against a field loss alone, it organises its predictions into source-centred wave packets aligned with the Voronoi partition of the sources, and the amplitude ablation rules out passive Green's-function reproduction. This is the primary scientific finding, and it answers SQ1: the map is learnable at every pair, with full-grid error rising gracefully from 0.240 to 0.339 across $16 \to 32$ through $64 \to 128$.

C2. Field accuracy is not solver usefulness. In the Dirichlet setting $c_k(r_0) = \lambda_k c_k(e_0)$ turns a field error of 0.084 into a true residual of $11.64\|b\|_2$, because the remaining error concentrates in the coherent modes the operator amplifies most. This answers SQ2 in the negative, and marks the preconditioned residual as the more faithful indicator of solver value.

C3. Warm-start improvement is real but modest. At $\beta = 0.3$ the full-grid warm start gives the lowest final true residual for all three pairs, a $5\times$ reduction at $16 \to 32$ that shrinks with frequency, with adaptive savings of 3.6%, 1.2%, and 0.5%. CSL handles the large eigenvalues; the learned operator influences the near-zero spectrum where CSL alone is weakest.

9.2. Future research

The directions below are in descending priority.

P1. Operator-aligned training. Replace the field-only loss with one penalising the operator-induced residual, or weighting modal components by eigenvalue magnitude: the lowest-cost change with the largest leverage, addressing C2 and C3 without a new architecture.

P2. Learned frequency-transfer preconditioning. Extend the one-time warm start to a learned multi-frequency cycle analogous to a multigrid V-cycle, in which learned restriction and prolongation act at every Krylov step, not only on the initial state.

P3. Heterogeneous media and the nonlinear regime. Learned transfer is most likely to help in the nonlinear frequency-domain problem, where no Green's-function superposition exists and a cross-frequency map can serve as a continuation mechanism for Newton-type solves; the C2 requirement carries over to each linearisation.

P4. Cost reporting and amortisation. Iteration count alone cannot judge acceleration; studies should report training solves, label and inference cost, and the break-even number of deployment solves.

Learned frequency transfer is solver-useful when its cross-frequency information enters in a form the operator and the Krylov solver can use.

Part IV

Appendix

10

Reproducibility record

This appendix records the configuration underlying the solver-facing claims. It is a compact run register rather than a second methods chapter. Within each reported warm-start comparison, the right-hand side, high-frequency FD/PML operator, CSL preconditioner, damping parameter, stopping rule, iteration budget, and FGMRES implementation are held fixed; only the initial guess is changed.

Table 10.1: Protocol lock for the main two-dimensional FD/PML experiments.

Item	Recorded configuration
Computational domain	512×512 FD/PML grid; PML width 112 on each side; 288×288 physical interior.
Frequency pairs	$16 \to 32$, $32 \to 64$, and $64 \to 128$.
Dataset	9600 exact full-grid FD/PML source–solution pairs per frequency pair, stored under `/orcd/pool/006/fkiewiet/freq2transfer/datasets_fdpml_2d/`; complex Gaussian sources with real and imaginary components retained.
Data partition	Per-pair training used a random train/validation split with validation fraction 0.2 and seed 42 (`train_flux_full_2d.py`); solver-facing right-hand sides are drawn from that per-pair held-out block. The pooled $70\%/15\%/15\%$ partition of Section 6.4.2 is the separate convention used for the hyperparameter sweep (`train_unet_hparam.py`).
Main learned model	Full-grid FD/PML `TransferUNet`, five levels, `base_ch=32`, trained to predict the full complex high-frequency field.
Fixed-budget evaluation	CSL-preconditioned FGMRES with $\beta = 0.3$, 40 iterations, and ten test right-hand sides per frequency pair.
Adaptive evaluation	True-relative-residual stopping tolerance 10^{-6}; 10, 3, and 1 test right-hand sides for $16 \to 32$, $32 \to 64$, and $64 \to 128$, respectively.
Preconditioning and outcome metrics	Left-inserted CSL preconditioner (Section 7.4.2); full-grid true residual $\|b - A_{\omega_H} x_k\|_2 / \|b\|_2$ as verification and stopping quantity, CSL-preconditioned residual as primary solver indicator.

Dataset generation, training, solver evaluation, and figure generation are retained as separate computational artefacts. For each thesis-quality result, the retained record consists of the launch or sweep specification, numerical metadata, selected checkpoint, summary table, residual curves, and figure-generation script. This makes a reported table entry traceable to both its numerical evaluation and the model checkpoint from which it was produced.

Earlier exploratory pipelines used different PML or data conventions. They are retained for diagnosis,

but are not combined with the final full-grid FD/PML comparison.

Table 10.2: Training-sample screening thresholds applied before use. Samples failing any criterion were discarded before training and before any solver-facing evaluation.

Screening criterion	Action
`NaN` or `Inf` in either field component	Sample discarded
Interior norm below 10^{-8} in either normalised field, source norm below 10^{-8}, or stored RMS below 10^{-12}	Sample discarded

Table 10.3: Run labels and checkpoints behind the body-text descriptions. The readable names used in Chapter 7 are listed against the literal run folders and checkpoints they refer to.

Body-text description	Literal run label / checkpoint
Analytic diagnostic CNN	`train4`; 29-channel analytic Green's-function model; checkpoint `model_N600.pt` under `results_train4/run_up_20260310_142852/checkpoints/`, used for the interpretability diagnostics of Section 7.3.
Amplitude-ablation run	`train6` ablation; `results_train6/run_up_20260313_011035`; phase-only training set with the $1/\sqrt{r}$ spreading factor removed.
Main 2D solver model	Full-grid FD/PML `TransferUNet`, `base_ch=32`, five levels; checkpoints `best.pt` under `pair_16_32_N9600_base32_L5_ep120_seed42`, `pair_32_64_N9600_base32_L5_ep120_seed42`, and `pair_64_128_N9600_base32_L5_ep120_seed42`, all within `freq2transfer/precond_2d_rigorous/flux_full_smoke/` on ORCD scratch.

10.1. Computational environment

The experiments ran across three environments chosen to match the cost profile of each stage: a local workstation, the Waveserver interactive machine, and the MIT ORCD Engaging cluster. The division of labour follows the workflow stages of Section 6.9 rather than any scientific variable.

10.1.1. Local workstation

A local machine running VS Code was used for code development, syntax checks, small one-dimensional validation runs, and figure generation from saved result files. No dataset generation or two-dimensional solver benchmark was run locally. This environment carries no queue and no wall-clock limit, which makes it the right place for the fast, interactive parts of the work and the wrong place for anything memory-bound.

10.1.2. Waveserver

The interactive machine `wave5b.mit.edu` was used for most solver and figure scripts, including the CSL–FGMRES warm-start evaluations and the one-dimensional diagnostics. It has 32 CPU cores and 377 GiB total RAM; at the time of the environment check, 125 GiB was available. As a shared interactive server it has no batch queue, so jobs start immediately, but it has no scheduler-enforced wall-clock limit either. The user memory limit was reported as unlimited by `ulimit -a`, so long probes such as the $64 \to 128$ adaptive run compete with other users for memory and cores and must be sized accordingly.

10.1.3. MIT ORCD Engaging cluster

The MIT ORCD Engaging cluster, a SLURM-scheduled HPC system, was used for the two-dimensional dataset generation and the GPU training jobs. The 9600-sample-per-pair datasets are stored at `/orcd/pool/`

006/fkiewiet/freq2transfer/datasets_fdpml_2d/, and the main full-grid FD/PML checkpoints are retained on ORCD scratch. Work on Engaging is submitted as batch jobs to partitions with enforced limits, and three constraints shaped the experimental design.

- *Queue waiting time.* GPU partitions are contended, so submitted jobs wait before starting. In the exported sacct record for completed sg2d_train jobs, queue waits ranged from 28 s to 4 h 21 min 32 s, with a median of 7 min 57 s.
- *Wall-clock limit.* The committed full-grid FD/PML GPU training entry point used the mit_preemptable partition with one GPU, 64 GB requested node memory, and a requested wall-clock time of $02{:}00{:}00$. The mit_preemptable partition has maximum wall-clock time 2–$00{:}00{:}00$. The mit_normal_gpu partition has default wall-clock time $02{:}00{:}00$ and maximum wall-clock time $06{:}00{:}00$.
- *Memory limit.* The full-grid FD/PML training script used batch_size=2 by default for the GPU training entry point. The sampled GPU allocation reported an NVIDIA L40S with 46068 MiB of GPU memory.

These limits explain several choices already made in the experimental design. Checkpoint selection on the best validation epoch (Section 6.8) is partly a response to the wall-clock cap, since a job that is pre-empted or times out can still be resumed from its last checkpoint. Retaining datasets and checkpoints in cluster storage rather than regenerating them avoids repeated queue waits. Splitting the work so that dataset generation and GPU training run on Engaging while the solver evaluations and figures run on the unqueued Waveserver keeps the latency-sensitive parts off the batch scheduler. The two-dimensional solver budget of forty FGMRES iterations and the ten test right-hand sides per pair were sized to fit inside these resource limits while still supporting the comparisons in Chapter 7.

11

Verification and result-selection checks

The following checks were completed before interpreting a learned warm start as solver-relevant. They test the numerical chain from data generation to solver evaluation; they do not turn the controlled FD/PML study into a claim of production-scale external validation.

Table 11.1: Verification checks supporting the reported comparisons.

Check	Evidence and consequence
Dataset audit	Each of the three $N = 9600$ FD/PML datasets passed its audit ($n_{\text{bad}} = 0$). The imaginary source component was retained, and the operator used to generate the data matched the operator used by the evaluator. The learning labels and solver residuals therefore refer to the same complex FD/PML problem.
Full-grid PML test	The PML is retained in both the FD/PML training target and residual evaluation. Raw and PML-zeroed initialisations are compared explicitly, so an accurate-looking interior field is not treated as sufficient evidence of operator compatibility.
Comparator lock	Within each warm-start comparison, only x_0 changes. The operator, right-hand side, preconditioner, damping, stopping criterion, and FGMRES implementation remain fixed.
Metric separation	Field error, true residual, and CSL-preconditioned residual are recorded separately. Field prediction quality is therefore not substituted for solution of the discrete Helmholtz equation.
Mechanism check	The 1D Dirichlet diagnostic tests the identity $c_k(r_0) = \lambda_k c_k(e_0)$ in the known sine eigenbasis, confirmed numerically up to rounding. This verifies the residual-amplification mechanism in a setting where it is analytically inspectable.
Scope check	Fixed-budget results use ten test right-hand sides per frequency pair. Adaptive convergence results use 10, 3, and one long probe, respectively; the latter is interpreted as a scale estimate rather than a statistical mean.

11.1. Result inclusion rule

The thesis distinguishes confirmatory results from exploratory development runs. Confirmatory solver claims use only experiments for which the data generator, high-frequency residual operator, PML convention, trained checkpoint, and evaluation protocol are mutually matched. Earlier exploratory runs with different data or right-hand-side conventions, including the `_repaired` and `RHSConfig` variants, are retained as development records but are not pooled with the final comparisons and do not support a thesis conclusion.

These checks support the internal validity of the stated FD/PML warm-start result. They do not establish robustness over alternative media, discretisations, random training seeds, or frequency ratios; those boundaries are stated explicitly in the discussion.

References

[1] Gang Bao et al. “Inverse scattering by a continuation method with initial guesses from a direct imaging algorithm”. In: *Journal of Computational Physics* 227.1 (2007), pp. 755–762. DOI: 10.1016/j.jcp.2007.08.020.

[2] Sean Molesky et al. “Inverse design in nanophotonics”. In: *Nature Photonics* 12 (2018), pp. 659–670. DOI: 10.1038/s41566-018-0246-9.

[3] Jean Virieux et al. “An overview of full-waveform inversion in exploration geophysics”. In: *Geophysics* 74.6 (2009), WCC1–WCC26. DOI: 10.1190/1.3238367.

[4] R. Gerhard Pratt et al. “Gauss–Newton and full Newton methods in frequency-space seismic waveform inversion”. In: *Geophysical Journal International* 133.2 (1998), pp. 341–362. DOI: 10.1046/j.1365-246X.1998.00498.x.

[5] Frank Ihlenburg. *Finite Element Analysis of Acoustic Scattering*. Vol. 132. Applied Mathematical Sciences. New York: Springer, 1998. DOI: 10.1007/b98828.

[6] Ivo M. Babuška et al. “Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?” In: *SIAM Journal on Numerical Analysis* 34.6 (1997), pp. 2392–2423. DOI: 10.1137/S0036142994269186.

[7] Oliver G. Ernst et al. “Why it is Difficult to Solve Helmholtz Problems with Classical Iterative Methods”. In: *Numerical Analysis of Multiscale Problems*. Ed. by Ivan G. Graham et al. Vol. 83. Lecture Notes in Computational Science and Engineering. Berlin, Heidelberg: Springer, 2012, pp. 325–363. DOI: 10.1007/978-3-642-22061-6_10.

[8] Lloyd N. Trefethen et al. *Spectra and Pseudospectra: the Behavior of Nonnormal Matrices and Operators*. Princeton, NJ: Princeton University Press, 2005.

[9] Yogi Ahmad Erlangga et al. “On a class of preconditioners for solving the Helmholtz equation”. In: *Applied Numerical Mathematics* 50.3–4 (2004), pp. 409–425. DOI: 10.1016/j.apnum.2004.01.009.

[10] Vandana Dwarka et al. *Stand-alone multigrid for Helmholtz revisited: towards convergence using standard components*. 2023. arXiv: 2308.13476 [math.NA]. URL: https://arxiv.org/abs/2308.13476.

[11] C. D. Riyanti et al. “A new iterative solver for the time-harmonic wave equation”. In: *Geophysics* 71.5 (2006), E57–E63. DOI: 10.1190/1.2231109.

[12] Laurent Demanet et al. “Scattering in flatland: efficient representations via wave atoms”. In: *Foundations of Computational Mathematics* 12.2 (2012), pp. 119–175. DOI: 10.1007/s10208-011-9106-2.

[13] Youcef Saad. “A flexible inner-outer preconditioned GMRES algorithm”. In: *SIAM Journal on Scientific Computing* 14.2 (1993), pp. 461–469. DOI: 10.1137/0914028.

[14] Zongyi Li et al. “Fourier Neural Operator for Parametric Partial Differential Equations”. In: *International Conference on Learning Representations*. 2021. URL: https://openreview.net/forum?id=c8P9NQVtmnO.

[15] Raphaël Pestourie et al. “Physics-enhanced deep surrogates for partial differential equations”. In: *Nature Machine Intelligence* 5 (2023), pp. 1458–1465. DOI: 10.1038/s42256-023-00761-y.

[16] Hongyu Sun et al. “Extrapolated Full-Waveform Inversion with Deep Learning”. In: *Geophysics* 85.3 (2020), R275–R288. DOI: 10.1190/geo2019-0195.1.

[17] Hugo Melchers et al. *Neural Green’s Operators for Parametric Partial Differential Equations*. 2024. arXiv: 2406.01857 [cs.LG]. URL: https://arxiv.org/abs/2406.01857.

[18] Alvin Bayliss et al. “On accuracy conditions for the numerical computation of waves”. In: *Journal of Computational Physics* 59.3 (1985), pp. 396–404. DOI: 10.1016/0021-9991(85)90119-6.

[19] Edip Baysal et al. “Reverse time migration”. In: *Geophysics* 48.11 (1983), pp. 1514–1524. DOI: 10.1190/1.1441434.

[20] Carey Bunks et al. “Multiscale seismic waveform inversion”. In: *Geophysics* 60.5 (1995), pp. 1457–1473. DOI: 10.1190/1.1443880.

[21] David Colton et al. *Inverse Acoustic and Electromagnetic Scattering Theory*. 2nd ed. Vol. 93. Applied Mathematical Sciences. Berlin, Heidelberg: Springer, 1998. DOI: 10.1007/978-3-662-03537-5.

[22] Jean-Pierre Bérenger. “A perfectly matched layer for the absorption of electromagnetic waves”. In: *Journal of Computational Physics* 114.2 (1994), pp. 185–200. DOI: 10.1006/jcph.1994.1159.

[23] Yogi Ahmad Erlangga et al. “A novel multigrid based preconditioner for heterogeneous Helmholtz problems”. In: *SIAM Journal on Scientific Computing* 27.4 (2006), pp. 1471–1492. DOI: 10.1137/040615195.

[24] Timothy A. Davis. “Algorithm 832: UMFPACK V4.3—An Unsymmetric-Pattern Multifrontal Method”. In: *ACM Transactions on Mathematical Software* 30.2 (2004), pp. 196–199.

[25] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics, 2003. DOI: 10.1137/1.9780898718003.

[26] Alan George et al. *Computer Solution of Large Sparse Positive Definite Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1981.

[27] Youcef Saad et al. “GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems”. In: *SIAM Journal on Scientific and Statistical Computing* 7.3 (1986), pp. 856–869. DOI: 10.1137/0907058.

[28] Martin B. van Gijzen et al. “Spectral Analysis of the Discrete Helmholtz Operator Preconditioned with a Shifted Laplacian”. In: *SIAM Journal on Scientific Computing* 29.5 (2007), pp. 1942–1958. DOI: 10.1137/060661491.

[29] Yogi Ahmad Erlangga et al. “Comparison of multigrid and incomplete LU shifted-Laplace preconditioners for the inhomogeneous Helmholtz equation”. In: *Applied Numerical Mathematics* 56.5 (2006), pp. 648–666. DOI: 10.1016/j.apnum.2005.04.039.

[30] Tuomas Airaksinen et al. “An algebraic multigrid based shifted-Laplacian preconditioner for the Helmholtz equation”. In: *Journal of Computational Physics* 226.1 (2007), pp. 1196–1210. DOI: 10.1016/j.jcp.2007.05.013.

[31] Björn Engquist et al. “Sweeping preconditioner for the Helmholtz equation: moving perfectly matched layers”. In: *Multiscale Modeling & Simulation* 9.2 (2011), pp. 686–710. DOI: 10.1137/100804644.

[32] Yael Azulay et al. “Multigrid-augmented deep learning preconditioners for the Helmholtz equation”. In: *SIAM Journal on Scientific Computing* 45.3 (2023), S127–S151. DOI: 10.1137/21M1433514.

[33] Bar Lerer et al. “Multigrid-augmented deep learning preconditioners for the Helmholtz equation using compact implicit layers”. In: *SIAM Journal on Scientific Computing* (2024). In press; preprint arXiv:2306.17486. Confirm volume/pages at proof stage. arXiv: 2306.17486 [cs.LG]. URL: https://arxiv.org/abs/2306.17486.

[34] Robert Geirhos et al. “Shortcut learning in deep neural networks”. In: *Nature Machine Intelligence* 2.11 (2020), pp. 665–673. DOI: 10.1038/s42256-020-00257-z.

[35] Michal Aharon et al. “K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation”. In: *IEEE Transactions on Signal Processing* 54.11 (2006), pp. 4311–4322. DOI: 10.1109/TSP.2006.881199.

[36] Yuhan Wu et al. *Are Deep Learning Based Hybrid PDE Solvers Reliable? Why Training Paradigms and Update Strategies Matter*. 2026. arXiv: 2602.06842 [math.NA]. URL: https://arxiv.org/abs/2602.06842.

[37] Leonardo Zepeda-Núñez et al. “The Method of Polarized Traces for the 2D Helmholtz Equation”. In: *Journal of Computational Physics* 308 (2016), pp. 347–388. DOI: 10.1016/j.jcp.2015.11.040.

[38] Michael L. Parks et al. “Recycling Krylov subspaces for sequences of linear systems”. In: *SIAM Journal on Scientific Computing* 28.5 (2006), pp. 1651–1674. DOI: 10.1137/040607277.

[39] Artem H. Sheikh et al. “Accelerating the Shifted Laplace Preconditioner for the Helmholtz Equation by Multilevel Deflation”. In: *Journal of Computational Physics* 322 (2016), pp. 473–490. DOI: 10.1016/j.jcp.2016.06.025.

[40] Nikola Kovachki et al. *Neural Operator: Learning Maps Between Function Spaces*. 2021. arXiv: 2108.08481 [cs.LG]. URL: https://arxiv.org/abs/2108.08481.

[41] Vandana Dwarka et al. “Scalable Convergence Using Two-Level Deflation Preconditioning for the Helmholtz Equation”. In: *SIAM Journal on Scientific Computing* 42.2 (2020), A901–A928. DOI: 10.1137/18M1192093.