Signal Processing Foundation for Interpretable ML

1. The Conceptual Scaffold

All of these experiments — scattering, Gabor filters, spectral flow, ablations — are built on classical signal processing ideas that originated long before deep learning. We're just reintroducing them in the context of interpretable machine learning.

Level 1: Signal as a Function

A histopathology tile $x(u, v)$ (or any image) is a 2-D signal:

$$x : \mathbb{R}^2 \to \mathbb{R}^3$$

(color image; three channels).

Our goal is to find transformations $\Phi(x)$ that:

- Are invariant to small shifts (translation invariance),
- Are stable to small deformations (Lipschitz continuous under warping),
- Retain discriminative information about underlying tissue structure.

These are exactly the design criteria of good representations in signal processing.

Level 2: Fourier Transform

Fourier analysis expresses a signal as a superposition of sinusoids:

$$x(u, v) = \int X(f_u, f_v) e^{j2\pi(f_u u + f_v v)} df_u df_v$$

- Gives global frequency content.
- Loses where those frequencies occur → no localization.

This is why pathology textures (nuclei, glands, fibers) can't be fully understood with Fourier alone.

Level 3: Windowed / Localized Transforms

We add localization in space → time-frequency analysis.

Two major developments:

Short-Time Fourier Transform (STFT):

- Multiply by a window $g(u, v)$ then Fourier transform.
- Trade-off: window size ↔ frequency resolution.

Wavelet Transform:

- Instead of fixed window, use scaling and translation of a mother wavelet ψ:

$$W_x(a, b) = \int x(t) \psi^*_{a,b}(t) dt$$

- Multiscale, more natural for images and textures.

These two — STFT and wavelets — are your foundation for understanding scattering and Gabor filters.

Level 4: Gabor Filters (Parametric Filterbanks)

A Gabor function is a Gaussian-modulated sinusoid:

$$g(u, v; f_0, \theta, \sigma) = \exp\left(-\frac{u'^2 + v'^2}{2\sigma^2}\right) \cos(2\pi f_0 u')$$

where

$$u' = u\cos\theta + v\sin\theta, \quad v' = -u\sin\theta + v\cos\theta$$

Parameters:

- f_0: central frequency
- θ: orientation
- σ: scale (bandwidth)

Interpretation: Each Gabor filter acts like a neuron tuned to a specific spatial frequency and orientation — much like cells in the human visual cortex (Daugman, 1985).

When we make these filters learnable, we're asking:

"Which frequency–orientation pairs best discriminate between tissue classes?"

Hence, we constrain f_0, σ etc. to stay interpretable.

Level 5: Scattering Transform

Proposed by Stéphane Mallat (2012), the scattering transform is a non-learned deep network built from wavelet convolutions and modulus operators.

$$S_0 x = x * \phi_J$$

$$S_1 x = |x * \psi_{\lambda_1}| * \phi_J$$

$$S_2 x = ||x * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \phi_J$$

$$\vdots$$

where:

- ψ_λ are bandpass wavelets (frequency + orientation),
- ϕ_J is a low-pass averaging filter (scale 2^J).

Interpretation:

- S_1 captures local energy per band (texture content),
- S_2 captures interactions between bands (e.g., pattern of edges).

It's provably:

- Translation-invariant
- Stable to deformations
- Linear in energy (Parseval-like conservation)

So when we compute scattering on PANDA tiles, we're creating a physically meaningful, mathematically grounded, and interpretable descriptor of tissue microtexture.

Level 6: From Filterbanks → Deep Nets

A CNN layer can be viewed as a learned filterbank.

Mathematically:

$$y_k(u,v) = \sigma(x * h_k(u,v) + b_k)$$

Each learned h_k is like a Gabor or wavelet tuned to a discriminative direction in frequency space. Thus, CNNs generalize classical DSP — but lose interpretability because filters are unconstrained.

Your parametric Gabor layer restores interpretability:

- Instead of random weights, filters are described by (f_0, θ, σ).
- You can visualize, constrain, and reason about them directly in frequency space.

Level 7: Quantitative Interpretability Metrics

All your "interpretability probes" have DSP meanings:

Probe	DSP Interpretation
Filter visualization	Impulse response & magnitude spectrum
Frequency-orientation histograms	Filterbank coverage of spectral plane
Filter importance	Energy or SNR per band
Band ablation	Spectral sensitivity / causal contribution
Spectral flow	Frequency migration during training
Spectral concentration score	Sparsity of discriminative energy
Robustness (stain/deform)	Lipschitz stability under perturbation

These are not arbitrary metrics — they map back to properties of stable, localized transforms in DSP theory.

Level 8: Physiology Alignment

In pathology, spatial frequencies correspond to biological scales:

Frequency band	Anatomical feature
Low (~0–0.1 cycles/µm)	Tissue layout, gland architecture
Mid (~0.2–0.4)	Nuclei density, cytoplasmic texture
High (>0.5)	Noise, fine fiber textures

So by analyzing which bands your model uses most, you can infer what physical structures it relies on — making interpretability physiologically meaningful.

---

2. What to Read (and in What Order)

Here's a focused reading plan that will help you build full intuition — in 5 layers, from basic DSP to advanced scattering theory.

🟦 Layer 1 – Classical Signal Processing

Build comfort with transforms, frequency domain, and filterbanks.

- **Alan V. Oppenheim & Ronald W. Schafer** – *Discrete-Time Signal Processing* (Ch. 4–8)
  - Fourier Transform, Sampling, and Digital Filters.
- **Mallat, S.** – *A Wavelet Tour of Signal Processing*, 3rd Edition.
  - Ch. 1–5: Continuous/discrete wavelets, multiresolution analysis.
- **Optional:** Gonzalez & Woods – *Digital Image Processing* (Ch. 4–6 on filtering and frequency domain).

🟩 Layer 2 – Wavelets & Time-Frequency Analysis

Connect to what "scales" and "orientations" mean in 2-D.

- **Mallat, S. (1999).** *A Wavelet Tour of Signal Processing* – Ch. 7–8 (2D wavelets, oriented wavelets).
- **Daugman, J. G. (1985).** "Uncertainty relation for space and spatial frequency as applied to Gabor filters." *JOSA A*.
  - The mathematical foundation of Gabor filters.
- **Addison, P. S. (2017).** *The Illustrated Wavelet Transform Handbook* – highly visual explanation.

🟨 Layer 3 – Scattering Transform

Understand the theoretical reason it's called "interpretable deep learning".

- **Mallat, S. (2012).** "Group invariant scattering." *Communications in Pure and Applied Mathematics*.
- **Bruna & Mallat (2013).** "Invariant scattering convolution networks." *IEEE TPAMI*.
- **Kymatio Documentation and Tutorials** (PyTorch library implementing scattering).

🟧 Layer 4 – Interpretability and Neural DSP

Where DSP meets modern ML and interpretability.

- **Andén & Mallat (2014).** "Deep scattering spectrum." *IEEE TASLP*.
- **Oyallon, Mallat, et al. (2018).** "Scattering networks for image classification." *ICLR*.
- **Raghu et al. (2017).** "SVCCA: Canonical Correlation Analysis of Representations in Neural Networks."
  - (for spectral flow / representation alignment ideas)

🟥 Layer 5 – Biomedical Signal/Image Analysis

Connect mathematical transforms to physical tissue structures.

- **Doyle et al. (2008).** "Wavelet-based analysis of histopathology images." *Medical Imaging*.
- **Komura & Ishikawa (2018).** "Machine Learning Methods for Histopathological Image Analysis." *Computational and Structural Biotechnology Journal*.
- **Tellez et al. (2019).** "Whole-slide image analysis in histopathology: A review." *Frontiers in Medicine*.

---

3. How to Study This Efficiently

Here's a good cycle to follow:

1. **Pick one concept per week** — e.g., Wavelets, then Gabor, then Scattering.
2. **Write the forward and inverse transform equations** in your own words.
3. **Visualize impulse responses and magnitude spectra** using small Python snippets.
4. **Relate each mathematical parameter** to an anatomical or perceptual meaning.
5. **Finally, revisit the notebook** and see where each equation lives in code.