Note on Semi-Empirical Theory of Learning and Weight Watcher

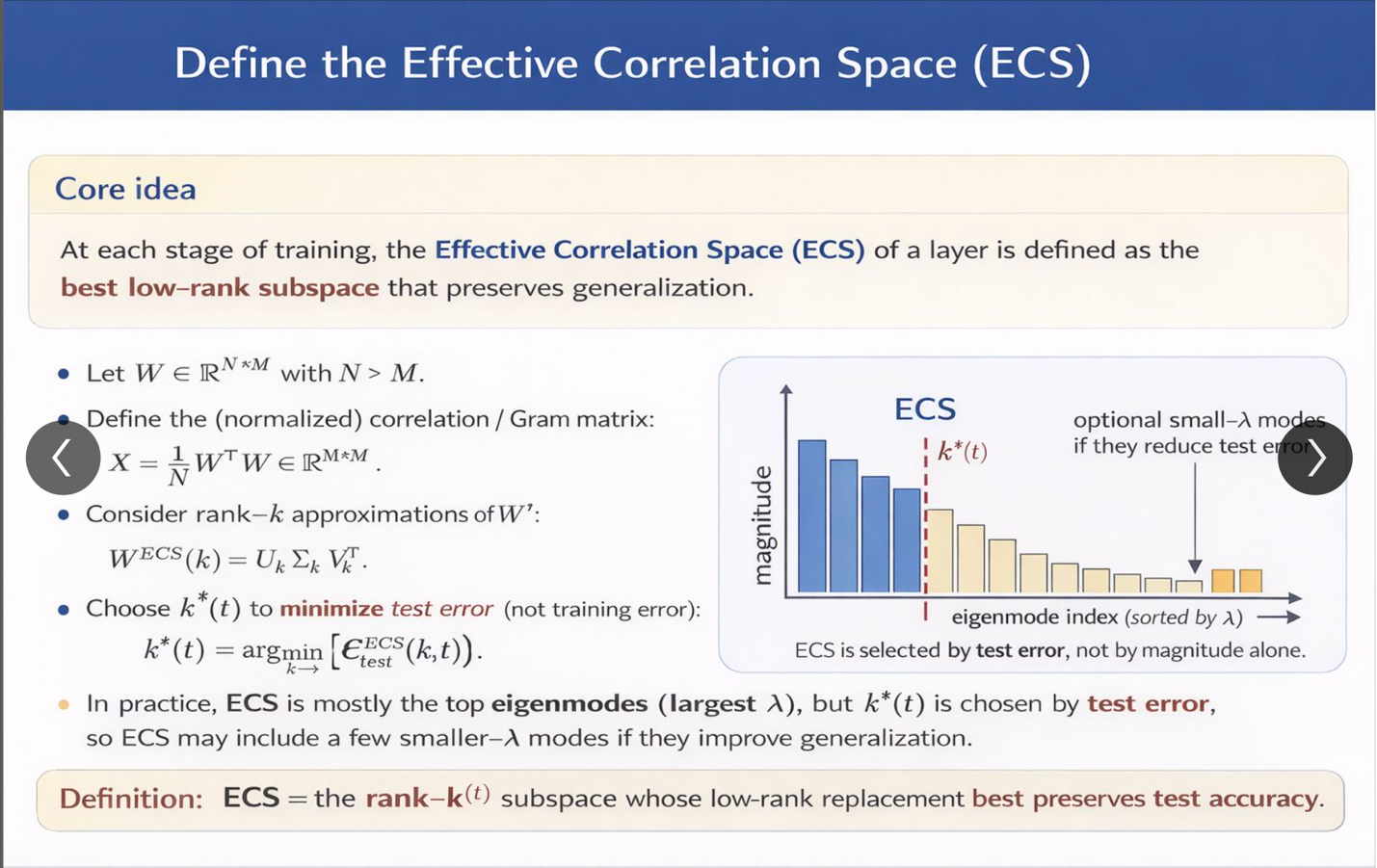
By Charles H Martin, 1/14/2026

WeightWatcher is based on the idea that neural-network (NN) learning behaves like a Renormalization Group (RG) flow. Here's some vibe coded slides to help explain what is going on  
  
In physics, RG describes how complex systems simplify as you zoom out. Irrelevant details fade away, and only the structures that matter at large scales remain.  
  
WeightWatcher applies this same intuition to neural networks.  
  
➤ 1️⃣ Not all parameters matter equally  
  
As a NN learns, most directions in weight space become irrelevant for generalization. What matters is a much smaller, structured subspace that actually controls performance.  
  
This is called the Effective Correlation Space (ECS).  
  
➤ 2️⃣ Learning looks like a flow, not just optimization  
Instead of thinking of training as “minimizing loss,” think of it as a flow:  
  
• important directions are amplified  
• redundant or noisy directions are suppressed  
• structure gradually emerges  
  
This mirrors how RG flows drive physical systems toward simpler, universal behavior.  
  
➤ 3️⃣ We can measure this flow directly  
WeightWatcher tracks layer metrics like the Trace Log (detX) condition and the power law exponent α that each quantify how correlated and structured a layer has become.  
  
• Intuitively, the TraceLog condition measures the effective volume of meaningful correlations in the layer.  
  
• As training progresses, the ECS shrinks, the TraceLog decreases, and, like a physical system, approaches a fixed point:

➤ 4️⃣ Generalization corresponds to a fixed point  
  
When a network generalizes well:  
• its effective structure becomes scale-invariant  
• correlations follow a universal pattern  
• a low-rank version of the layer reproduces test accuracy  
  
In RG language, the network has reached a non-trivial fixed point where   
  
• the ERG TraceLog condition vanishes,   
 = ∑ ln λ = 0, λ ∈ ECS  
  
• the weightwatcher power law exponent α = 2,   
 ρ(λ) ∼ λ^(-α), α = 2, λ ∈ ECS  
  
And, as shown in the SETOL paper, the layer weight matrix can be replaced with its ECS exactly.

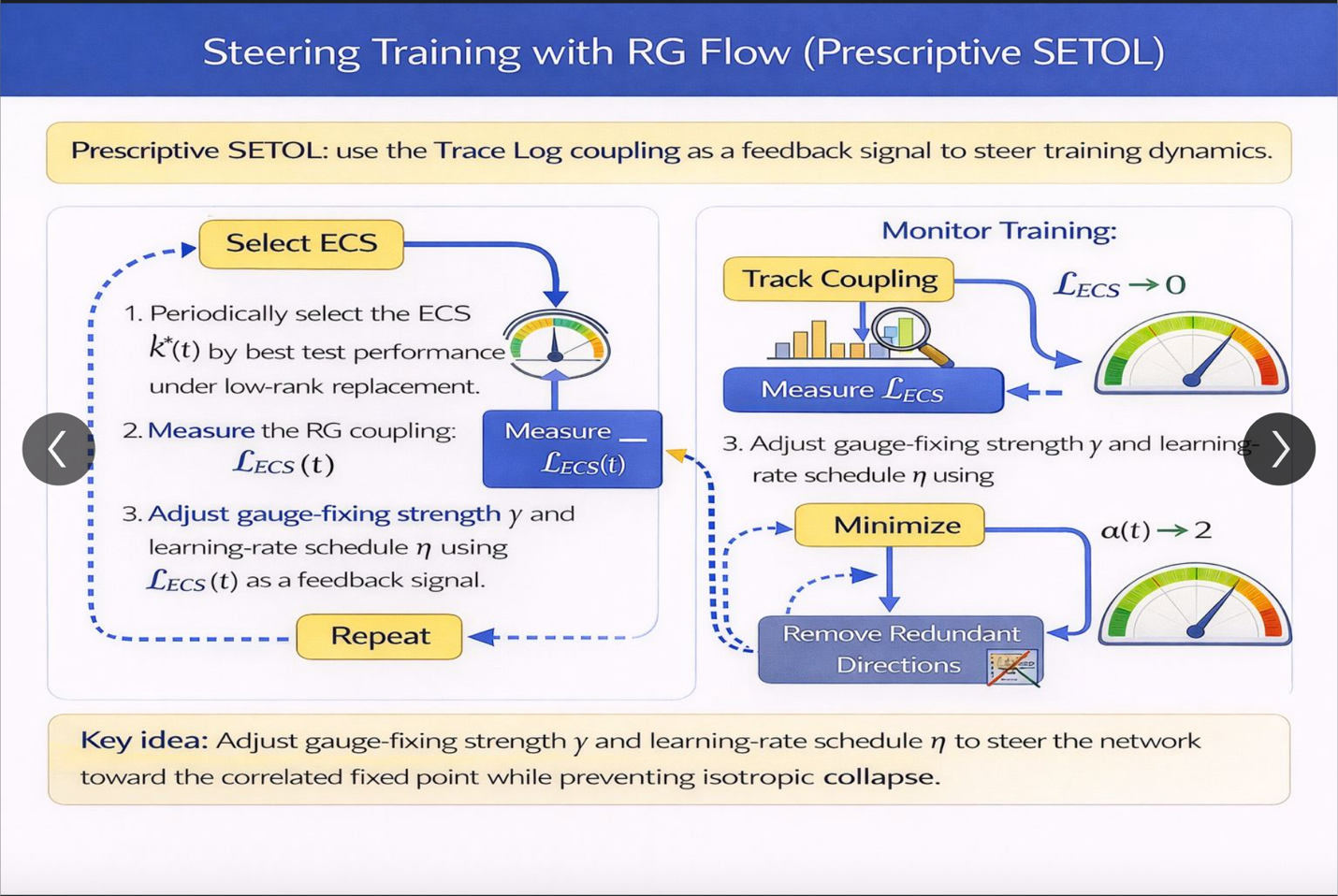
A screenshot of a computer

AI-generated content may be incorrect.



A screenshot of a computer

AI-generated content may be incorrect.



A screenshot of a computer

AI-generated content may be incorrect.

# References

[1] [Charles H. Martin’s post on Linkedin , 1/14/2026](https://www.linkedin.com/posts/charlesmartin14_weightwatcher-is-based-on-the-idea-that-neural-network-ugcPost-7417293628475232256-aKHw?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAFZfUoBgPoGUucdnvtwuzPv79P8VHj6uvk)

[2] [SETOL: A Semi-Empirical Theory of Deep Learning, Charles H. Martin and Christopher Hinrichs, 2025](https://github.com/dimitarpg13/geometric_deep_learning/blob/main/articles/SETOL/SETOL-A_Semi-Empirical_Theory_of_Deep_Learning_Martin_2025.pdf)

[3] WeightWatcher: Data-Free Diagnostics for Deep Learning: <https://weightwatcher.ai/>

[4] [WeightWatcher, HTSR theory, and the Renormalization Group, calculated content blog, December 24, 2024](https://calculatedcontent.com/2024/12/24/weightwatcher-htsr-theory-and-the-renormalization-group/)