Stage 11 — Action Plan for Fresh Thread

This action plan summarizes the key directions from the current thread and sets up a clear roadmap for continuing work on Stage 11 in a new thread. The focus is to stabilize the well, diagnose and suppress phantom wells, and prepare for Stage 12 benchmarking.

1. Immediate Priorities

• Run Step-5 baseline (S5-3) with --dump_surfaces_dir to capture per-sample Eraw/Eperp surfaces.
• Use visualization tools to produce surfaces and ridge plots.
• Label and analyze wells (TRUE vs PHANTOM) for 10–20 samples to directly visualize phantom problem.

2. Diagnostic Sweeps

• Phantom Index (PI): count local minima not aligned to truth; track across gates/ablations.
• Prototype ablations: remove hinge/deriv/half variants one at a time; measure hallucination deltas.
• Aggregation test: confirm softmin+consensus vs. median.
• Orthogonalization toggle: flip_v ortho on vs. off.

3. Interventions to Test

• Consensus tighten (k↑, eps↓) for flip_v.
• Gate adjustments: raise raw_floor_v, lower residual_ceiling_v.
• Prototype pruning: drop shapes correlated with phantom wells.
• Inhibition tuning: gentler inhibition to reduce ripple-induced dents.

4. Success Criteria

• Recall = 1.0 (no omissions).
• Hallucination ≤ 0.26 (≥10% drop from ~0.29).
• flip_v hallucination ≤ 0.24.
• Margins healthy (mean ≥1.6, min ≥0.9).
• Grid similarity/accuracy not degraded vs. S5-3 baseline.

5. Preparing for Stage 12

• Lock in a stable Stage-11 config once phantom suppression is validated.
• Define seeds, sample sizes, regimes for Stage-12 benchmarking (2,000+ samples/condition).
• Automate comparison against stock with bootstrap CIs for F1/hallucination.
• Establish reporting pipeline: tables, delta plots, and PDF dossiers.

Conclusion

The well is the path to suppressing hallucinations, but phantom wells remain. Stage 11 extension work will focus on diagnosing and eliminating phantom attractors. This plan prepares for a fresh thread, leading into Stage 12's rigorous benchmarking.