

IGT - Information Gravity Theory - Part 2 (IGT2)

Appendix (execution logs)

Selection from RTX 5090 execution sessions, June 2026. Contains successful, complete runs that support the observation in §4.2C. Environmental errors (gated access, authentication, architecture incompatibilities) have been omitted as irrelevant to the result.

Environment

torch 2.11.0+cu128

transformers 5.9.0

bitsandbytes 0.49.2

Hardware: RTX 5090 – 24Gb VRAM, 4-bit quantization (Unsloth bnb-4bit), Float16 calculation

Run 1: Qwen-2.5-14B-Instruct (layer 20), full oscillation monitoring

Prompt: requesting a proof for Goldbach 's Conjecture (unsolved problem, L3 complexity).

Complete, consistent, non-stopping, pure profiling of V_drift consecutively on each token:

Hook successfully recorded on layer 20.

[Prefill complete. Step-by-step monitoring enabled.]

Pro	V_drift:	0.0000 (anchor)
ble	V_drift:	0.4637
ma	V_drift:	0.4397
pe	V_drift:	0.2702 ← drop (function word)
care	V_drift:	0.5784 ← peak (content word)
ai	V_drift:	0.4778
men	V_drift:	0.6095
ț	V_drift:	0.5558
ion	V_drift:	0.5388
at	V_drift:	0.5141
-o	V_drift:	0.2995 ← drop
este	V_drift:	0.4370
c	V_drift:	0.4831
unos	V_drift:	0.4867
cut	V_drift:	0.6368 ← peak (content boundary)
ă	V_drift:	0.5350
sub	V_drift:	0.2119 ← drop (preposition)
num	V_drift:	0.4438
ele	V_drift:	0.4383
de	V_drift:	0.4268
Con	V_drift:	0.4585
ject	V_drift:	0.4582
ura	V_drift:	0.4146
lui	V_drift:	0.3116 ← drop (function word)
Gold	V_drift:	0.5142
bach	V_drift:	0.6005 ← peak (proper noun / content)
.	V_drift:	0.5333
Ace	V_drift:	0.6345 ← peak
asta	V_drift:	0.6218
afirm	V_drift:	0.5002
ă	V_drift:	0.6144
că	V_drift:	0.4184
or	V_drift:	0.3290 ← drop
ice	V_drift:	0.3763

num	V_drift:	0.2547	← drop
ăr	V_drift:	0.5168	
par	V_drift:	0.3282	
mai	V_drift:	0.4795	
mare	V_drift:	0.5669	
sau	V_drift:	0.3680	
egal	V_drift:	0.4889	
cu	V_drift:	0.4373	
4	V_drift:	0.4605	
Poate	V_drift:	0.4088	
fi	V_drift:	0.5012	
sc	V_drift:	0.3583	
ris	V_drift:	0.4915	
ca	V_drift:	0.5230	
suma	V_drift:	0.2997	← drop
a	V_drift:	0.3337	
dou	V_drift:	0.4557	
ă	V_drift:	0.5205	
num	V_drift:	0.3132	← drop
ere	V_drift:	0.5212	
prime	V_drift:	0.4252	
.	V_drift:	0.4040	
De	V_drift:	0.5813	
a	V_drift:	0.6489	← peak
sem	V_drift:	0.5622	
ene	V_drift:	0.5194	
a	V_drift:	0.6026	
,	V_drift:	0.6117	
exist	V_drift:	0.4130	
ă	V_drift:	0.6197	← peak
o	V_drift:	0.2936	← drop (function word)
variant	V_drift:	0.2961	
ă	V_drift:	0.5373	
mai	V_drift:	0.2635	← drop
put	V_drift:	0.4937	
ern	V_drift:	0.4721	
ic	V_drift:	0.4357	
ă	V_drift:	0.5018	
a	V_drift:	0.6192	← peak
ace	V_drift:	0.3951	
ste	V_drift:	0.5386	
i	V_drift:	0.3498	
conject	V_drift:	0.3396	
uri	V_drift:	0.6402	← peak
care	V_drift:	0.3891	

[FINAL LOCAL RESULT - coherent, correct text]:

The problem you mentioned is known as Goldbach's Conjecture. It states that any even number greater than or equal to 4 can be written as the sum of two prime numbers. There is also a stronger version of this conjecture that [...]

[RO] "Problema pe care ai menționat-o este cunoscută sub numele de Conjectura lui Goldbach. Aceasta afirmă că orice număr par mai mare sau egal cu 4 poate fi scris ca suma a două numere prime. De asemenea, există o variantă mai puternică a acestei conjecturi care [...]"

Observation: band 0.21-0.65, valleys on functional words ("pe", "-o", "sub", "lui", "or", "num", "suma", "o", "mai"), peaks on content ("care", " cut "/known, " bach "/ Goldbach, "Ace"/This, "uri"/conjectures). The generated text is coherent and correct throughout, and the oscillation does NOT signal an error, but is a feature of normal generation.

Control run: Qwen-2.5-14B (layer 20), V_drift + L2 norm ||h||

Negative control test: the same run, with the L2 norm of the hidden state measured simultaneously with V_drift, to exclude the hypothesis that the **directional oscillation** is an artifact of the *massive phenomenon activations* (extreme magnitude of activation). Representative excerpt (peaks and valleys of drift, with the norm attached):

token	V_drift	h	
pe	0.2702	82.7	← drift valley / average norm
care	0.5784	82.1	← drift peak / average norm (high drift, unchanged norm)
men	0.6095	87.1	← drift peak / average norm
cut	0.6368	97.4	← drift peak / average norm
sub	0.2119	87.0	← drift valley / average norm (low drift, unchanged norm)
ura	0.4146	113.8	← medium drift / high norm
Gold	0.5142	106.7	←
.	0.5333	141.2	← medium drift / MAXIMUM norm (does not coincide with drift peak)
Ace	0.6345	75.6	← drift PEAK / LOW norm (clear decoupling)
suma	0.2997	94.7	← drift valley / high norm (anticorrelation)
a	0.6489	71.0	← drift peak / low norm
ă	0.6197	71.0	← drift peak / low norm
o	0.2936	76.0	← drift valley / low norm
uri	0.6402	104.7	← drift peak / high norm

Analysis: total range of the norm ≈ 68 -141 (max /min ratio $\approx 2\times$), without spikes of the order of 10^3 - 10^4 that would indicate massive activations. Norm and V_drift are decoupled: the "Ace" drift peak (0.63) has a low norm (75.6), while the "sum" wave (0.30) has a high norm (94.7); the maximum norm (141.2, per point) coincides with the average drift, not with the peak. Conclusion: the directional oscillation (cosine, magnitude invariant) is a phenomenon distinct from the magnitude of the activation. The confound massive-activations is excluded.

Run 2: Gemma-2-9B-it (layer 21), pattern reproduction + norm check

Reproduction on the second architecture, on the correct identity layer (21 for Gemma-2-9B), with V_drift and L2 norm measured simultaneously. Representative excerpt:

token	V_drift	h	
Nu	0.0000	353.7	(anchor)
este	0.2225	264.7	
adev	0.1238	228.2	
ărat	0.2472	247.4	
ăr	0.4387	274.9	← drift peak / average norm
reg	0.4321	289.9	
suma	0.1585	216.9	← drift valley / average norm
două	0.3941	195.6	
Gold	0.2294	207.9	← content, but small drift (structural exception)
bach	0.2715	222.2	
,	0.4276	324.7	← high drift / HIGH NORM
și	0.3823	337.6	← functional, but high drift (structural exception)
demonstra	0.1043	233.9	← minimum drift / average norm
I	0.3791	341.8	← medium drift/MAXIMUM norm (does not coincide with drift peak)
sunt	0.2279	328.2	

Analysis: norm band $\approx 180\text{-}342$ (max / min ratio $\approx 1.9\times$), completely different from Qwen ($\sim 68\text{-}141$), but without massive spikes activation. The norm- drift decoupling is reproduced: the maximum norm (341.8, "I") falls on the average drift ; the peak of the " ár " drift (0.44) falls on the average norm. The valley/peak structure on the token type is noisier than on Qwen (see "and" functional with large drift, " Goldbach " content with small drift). The generated text is correct ("It is not true that any number... Goldbach 's conjecture... still unproven"). The drift band Gemma ($\sim 0.10\text{-}0.45$) is lower than Qwen ($\sim 0.21\text{-}0.65$), confirming that the band is model-specific.

[FINAL LOCAL RESULT, coherent, correct text]:

It is not true that any integer greater than 2 can be written as the sum of two prime numbers. This statement is known as Goldbach 's conjecture, and is still unproven. [...]

[RO] "Nu este adevărat că orice număr întreg mai mare decât 2 poate fi scris ca suma a două numere prime. Această afirmație este cunoscută sub numele de ipoteza lui Goldbach, și este încă nedemonstrată. [...]"

Note: for Gemma, the answer "It is not true that every integer..." is correct - Goldbach is referring to **even numbers**; the model correctly signals that the formulation with "integer" (which includes odd numbers) is false.

Reproducibility notes

- The nominal band is model specific (Qwen $\sim 0.21\text{-}0.65$); it is not a universal constant.
- The measurement layer matters (intermediate: 20/21 out of $\sim 40/42$).
- 4-bit quantization does not suppress the phenomenon.
- The measurement is V_{drift} *consecutive* (token -to- token), not drifting from the prefill anchor.