Linear Layers Stacked Any Number Deep Are Exactly One Layer, and One Bend Between Them Doubles the Pieces at Every Layer

— Three linear layers and their product agree on 200 inputs to a largest relative difference of 5.192×10^{-15}, so depth does nothing — the separator is whether a nonlinearity such as ReLU sits between the layers — [Paper 745]

Yuuki YAMAGISHI (山岸 勇輝)
Yamagishi Complexity Torus Lab, Japan
September 2026

— From thinking to syncing. —

Abstract

Deeper networks are said to represent more. That is true only when something bends between the layers.The cases where depth is worth nothing and where it is worth an exponential are set side by side with the same parts.No new theorem or law is claimed. First, linear layers stacked are one layer——three layers against their product: a largest relative difference of 5.192×10^{-15} (Section 2). **Second, what a line cannot make, depth cannot make——the best line fit to $|x|$ leaves a mean squared error of 0.083342** (Section 2). **Third, two bends make it exactly——two ReLUs give $|x|$ with error 0** (Section 2). **Fourth, and this is the core. With bends, depth acts exponentially——composing the tent map L times gives 2^L pieces, 1024 at $L = 10$** (Section 3). **Fifth, a shallow build needs units by orders——the same 1024 pieces take 20 ReLUs in depth and 1023 in one layer, a factor of 51.150000** (Section 3). **Sixth, the separator is whether a nonlinearity such as ReLU sits between the layers——the value of depth splits into zero or exponential, with nothing between** (Section 4). **Keywords**: activation function, ReLU, composition of linear maps, piecewise linear, depth and expressiveness, presence of a nonlinearity.

> **Scope of this paper (scope note)**: **No new theorem or law is claimed——that compositions of linear maps are linear, the counting of pieces in piecewise-linear networks, and the difference depth makes to expressiveness are all known (Telgarsky 2016, Goodfellow 2016).Learning is not treated——how weights are found is not treated; only representable functions are counted.One dimension only——inputs are one-dimensional and expressiveness is counted as pieces.No real network is claimed——particular networks and tasks are not treated.The shallow count is limited in meaning——the units needed in one layer are those for building a one-dimensional piecewise-linear function.Relation to earlier papers**: **Paper 330 showed that the squaring of the condition number in the normal equations is an identity, not an approximation——there a product of matrices worsened conditioning, as an identity. Here stacking linear layers yields one layer, also as an identity, and the product erases depth. What a product of matrices increases, and what it erases.**What is added is **confirming numerically that three linear layers equal one**, **showing that a line leaves $1/12$ on $|x|$**, **counting 2^L pieces for the composed tent map up to $L = 10$**, **showing that a shallow build needs $\mathbf{51.150000}$ times the units**, and **placing the separator on the presence of a nonlinearity**.

1. Notation and the Question

Setting: **let each layer be a matrix W_k. Without a nonlinearity, stacked layers are a product of matrices. ReLU is $\max(0, x)$ and makes one bend.**

$$W_3 W_2 W_1 x = (W_3 W_2 W_1)\, x, \qquad |x| = \mathrm{ReLU}(x) + \mathrm{ReLU}(-x), \qquad t(x) = 2\,\mathrm{ReLU}(x) - 4\,\mathrm{ReLU}\big(x - \tfrac{1}{2}\big)$$

t is the tent map taking $[0, 1]$ to $[0, 1]$, written with two ReLUs.

Question: **what does making a network deeper increase?**

2. Linear, Depth Does Nothing

Computation 2.1: **three random 6×6 matrices were stacked, and 200 inputs were passed through them and through their product.**

quantity	value
largest relative difference, three layers against one	5.192×10^{-15}
mean squared error of the best line fit to $\lvert x \rvert$	0.083342
difference from $1/12$	8.335×10^{-6}
error of $\lvert x \rvert$ built from two ReLUs	0

Observation 2.2

three linear layers are exactly one layer.The outputs differ only by rounding, 5.192×10^{-15}——they are equal not approximately but by the identity of the matrix product.

Observation 2.3

so what a line cannot make, depth cannot make.The best line on $[-1, 1]$, the constant 0.5, leaves 0.083342, matching $1/12$——no number of layers removes any of it.

Observation 2.4

two bends bring the error to 0.Because $|x| = \mathrm{ReLU}(x) + \mathrm{ReLU}(-x)$——what increased the representable was the bend, not the layer count.

3. With Bends, Depth Acts Exponentially

Computation 3.1: the tent map was composed L times, and its linear pieces on $[0, 1]$ counted by changes in the direction of slope.

L	ReLUs in depth	pieces	ReLUs needed in one layer
1	2	2	1
2	4	4	3
3	6	8	7
4	8	16	15
6	12	64	63
8	16	256	255
10	20	1024	1023

Observation 3.2

this is the core: the pieces number exactly 2^L.Each further composition folds every existing piece into two ——the same two ReLUs add nothing when linear and double the pieces when composed.

Observation 3.3

a shallow build needs units by orders.In one dimension a function of 1024 pieces needs one ReLU per bend in a single layer, 1023, against 20 in depth, a factor of 51.150000——what depth earns is an exponent in the number of bends.

Observation 3.4

the value of depth does not grow continuously.Zero without nonlinearity, exponential with it, and no form between——“a little nonlinear, a little gain” does not hold in the count of pieces.

4. The Separator

Observation 4.1

the separator is whether a nonlinearity such as ReLU sits between the layers.Without it depth is worth zero; with it, an exponential——“deeper is stronger” has no truth value until this condition is stated.

Observation 4.2

beside Paper 330.Paper 330 showed that the condition number’s squaring in the normal equations is an identity, not an approximation——there the product of matrices worsened conditioning; here it erases depth. Both are identities, the two sides of what a product increases and erases.

Observation 4.3

what is not said.Learning, multidimensional inputs and real networks are not treated; only pieces are counted——what is shown is that the effect of depth splits into zero or exponential on the presence of a nonlinearity.

5. Roll Call and Closing

1. **three against one:** $\mathbf{5.192 \times 10^{-15}}$ (Computation 2.1): **an identity.**
2. **line error** $\mathbf{0.083342}$ (Computation 2.1): **depth does not reduce it.**
3. **from** $\mathbf{1/12}$ **by** $\mathbf{8.335 \times 10^{-6}}$ (Observation 2.3): **matching the closed value.**
4. **two ReLUs, error** $\mathbf{0}$ (Observation 2.4): **the bend makes it.**
5. **pieces** $\mathbf{2^L}$ (Computation 3.1): **the core.**
6. **1024 pieces at** $\mathbf{L = 10}$ (Computation 3.1): **20 ReLUs.**
7. **1023 in one layer** (Computation 3.1): **by orders.**
8. **a factor of** $\mathbf{51.150000}$ (Observation 3.3): **what depth earns.**
9. **zero or exponential** (Observation 3.4): **nothing between.**
10. **a nonlinearity between layers** (Observation 4.1): **the separator; Paper 330.**

Closing: **Linear layers stacked any number deep are exactly one layer, and one bend between them doubles the pieces at every layer.Three linear layers and their product differ on 200 inputs by at most** $\mathbf{5.192 \times 10^{-15}}$**, equal by the identity of the matrix product——so the** $\mathbf{0.083342}$ $\mathbf{(1/12)}$ **left by the best line fit to** $\mathbf{|x|}$ **is not reduced by any number of layers.Two ReLUs make** $\mathbf{|x|}$ **exactly, so what increased the representable was the bend, not the layer count.And composing a two-ReLU tent map** $\mathbf{L}$ **times gives exactly** $\mathbf{2^L}$ **pieces, 1024 at** $\mathbf{L = 10}$**: 20 ReLUs in depth against 1023 in one layer, a factor of 51.150000——the value of depth is zero without nonlinearity and exponential with it, with no form between.The separator is whether a nonlinearity such as ReLU sits between the layers.Placed among the earlier papers——in Paper 330 a product of matrices worsened conditioning as an identity; here it erases depth as an identity.To be honest——learning, multidimensional inputs and real networks are not treated; only pieces are counted.**

References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

2. Telgarsky, M. (2016). Benefits of depth in neural networks. *Conference on Learning Theory*.

3. Yamagishi, Y. (2026). *That the Condition Number Squares Is an Identity, Not an Approximation* [Paper 330]. Zenodo. 10.5281/zenodo.22681423

— From thinking to syncing. —

Yuuki YAMAGISHI (山岸 勇輝)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)