nature portfolio

https://doi.org/10.1038/s41586-022-05340-6

Supplementary information

Dendrocentric learning for synthetic intelligence

In the format provided by the authors and unedited

SUPPLEMENTARY INFORMATION FOR

"DENDROCENTRIC LEARNING FOR SYNTHETIC INTELLIGENCE"

KWABENA BOAHEN

I ENERGY PER INFERENCE FOR GPT-3 ON PIXEL-4

In this section, I estimate how much energy it would take to run GPT-3 on a Pixel 4 phone. To run 109M-parameter BERT on 128 tokens, Pixel 4 takes 0.34 seconds[1]. Extrapolating linearly, running 175B-parameter GPT-3 on 2×2,048 tokens (1.4 tokens/word) will take 290 minutes. This rate of 10.1 wpm is 14.9-fold slower than average conversation rate (150 wpm). Thus, Pixel 4's processors need to operate 14.9-fold faster. But consuming electricity at 14.9-fold its 10W peak-rate will deplete Pixel 4's 10.4 Whr battery (2.8 Ah × 3.7V) in 4.2 minutes. To converse all day (12hr), it needs a 172-fold larger battery.

---

[1] Sun, Z. et al. (2020). MobileBERT: a compact task-agnostic BERT for resource-limited devices. *arXiv preprint* arXiv:2004.02984.

II WIRING ANALYSIS

In this section, I derive power laws that scale energy use with number of neuronal units in two and three-dimensional chips when routing is circuit-switched and compare this approach with packet-switched routing.

Assumptions and Definitions

Circuit switching, developed for telephone networks, assigns each unit its own wire and signals once on it. I make these three generic assumptions:

1. Wires occupy most of the space.
2. Propagating signals across wires dominates energy use.
3. A signal uses energy proportional to the distance it travels.

Unlike the analysis presented in the text (Box 1), I do not make reference to a specific architecture[2]. In lieu, I make use of these two operational definitions:

1. A *projection* routes a unit's output globally using $W^{1/D}$ units of wire.
2. A *connection* makes one synapse locally using one unit of wire.

Here $W = M(W^{1/D} + n)$ is the total amount of wire used to layout a deep net with M units—each with one projection and n connections—in D dimensions.

For example, consider a deep net with 256 units, each with one projection and 80 connections. Substituting 256 for M and 80 for n in the equation above and solving for W yields 320^2 for $D = 2$ and 30.5^3 for $D = 3$. Hence, the neural net occupies a square with sides 320 units long or a cube with sides 30.5 units long. The square's 102,400 square-unit area and the cube's 28,280 cubic-unit volume also give the total amount of wire used, in units of a wiring-pitch-sided pixel ($D = 2$) or voxel ($D = 3$). Thus, under my assumptions, the square would consume 3.6-fold more energy per inference.

How to Keep Wire per Synapse Constant

Amortizing a projection's cost across its unit's connections scales wire linearly with the total number of synapses (constant wire per synapse)[3]. To realize this scaling, connections per unit n must be proportional to projection length $W^{1/D}$ as wire per synapse is given by

$$w \equiv \frac{W/M}{n} = \frac{W^{1/D}}{n} + 1 \tag{1}$$

substituting the expression for W given above. Scaling n as $W^{1/D}$ keeps $W^{1/D}/n$ constant, and hence wire per unit W/M scales as n, or as $W^{1/D}$. Conversely, W scales as $M^{D/(D-1)}$ and n scales as $M^{1/(D-1)}$. Hence, n and W/M both scale as $M^{1/(D-1)}$—or as M in 2-D ($D = 2$) and as $\sqrt{M}$ in 3-D ($D = 3$). Thus, if the fraction of units that signal does not change with M, then energy per inference scales as $(W/M)M = M^{D/(D-1)}$—or as M^2 and $M^{3/2}$—the number of synapses in 2-D and 3-D, respectively.

Scaling connections per unit n as $M^{1/(D-1)}$ to keep wire per synapse w constant supports full layer-to-layer connectivity ($n \times n$) from $M^{1/(D-1)}$ units to $M^{1/(D-1)}$ units—M-to-M in 2-D and $\sqrt{M}$ -to-$\sqrt{M}$ in 3-D (as assumed in Box 1). Dividing M by $M^{1/(D-1)}$ yields the number of fully connected layers: $l = M^{(D-2)/(D-1)}$. This number is constant in 2-D and proportional to $\sqrt{M}$ in 3-D. Were connections per unit n to scale slower than $M^{1/(D-1)}$, wire per synapse w would increase

with neuron count, and thus energy per inference would increase even more rapidly with synapse count.

Role of Pre-Exponential Factors that Scale Width and Depth

To explore the role of pre-exponential factors, I substitute $n=(\gamma/\kappa)M^{1/(D-1)}$ units per layer into Eq. 1 and obtain

$$\frac{W}{Mn}=1+\frac{\kappa}{\gamma}\frac{W^{1/D}}{M^{1/(D-1)}}=1+\left(\frac{\kappa}{\gamma}\right)^{1-1/D}\left(\frac{W}{Mn}\right)^{1/D}\iff w=1+\left(\frac{\kappa}{\gamma}\right)^{1-1/D}w^{1/D} \tag{2}$$

Increasing γ expands width n and shrinks depth $l=(\kappa/\gamma)M^{(D-2)/(D-1)}$; κ allows geometrical differences between 2-D and 3-D to be accounted for. Solving this equation yields the following expressions for wire per synapse

$$w=1+\frac{1+\sqrt{1+2\gamma}}{\gamma}\text{ and }w=1+\frac{3}{2\gamma}\left(\rho(\gamma)+\frac{1}{\rho(\gamma)}\right)\text{ with }\rho(\gamma)=\sqrt[3]{\gamma+\sqrt{\gamma^2-1}} \tag{3}$$

in 2-D ($\kappa=2$) and 3-D ($\kappa=3^{3/2}/2$, $\gamma\geq 1$), respectively. In both cases, w is indeed constant. Thus, the total amount of wire, $W=wMn$, scales as $(\gamma/\kappa)M^{D/(D-1)}$, and so does energy per inference when the fraction of units that signal is constant. This result confirms that energy per inference scales quadratically with unit count in 2-D and with a power of three-halves in 3-D.

I reanalyze the 2-D case with n scaling not as M, but rather as $\sqrt{M}$, so that l scales similarly (as in Box 1). Substituting $D=2$ and $n=(\gamma/\kappa)\sqrt{M}$ in Eq. 1 yields

$$\frac{W}{Mn}=1+\frac{\kappa}{\gamma}\left(\frac{W}{M}\right)^{1/2}=1+M^{1/4}\left(\frac{\kappa}{\gamma}\right)^{1/2}\left(\frac{W}{Mn}\right)^{1/2}\iff w=1+M^{1/4}\left(\frac{\kappa}{\gamma}\right)^{1/2}w^{1/2}$$

This equation is equivalent to Eq. 2 with γ in that equation replaced by $\gamma/\sqrt{M}$. Hence, I adapt Eq. 2's solution (Eq. 3 *left*) to obtain the following expressions for wire per synapse and total wire

$$w=1+\frac{1+\sqrt{1+2\gamma/\sqrt{M}}}{\gamma/\sqrt{M}}\text{ and }W=\frac{M^2}{2}\left(1+\gamma/\sqrt{M}+\sqrt{1+2\gamma/\sqrt{M}}\right)$$

For $\sqrt{M}\gg\gamma$, $W\approx M^2/2$, and hence energy per inference also scales quadratically with unit count in this 2-D case. This result proves that replacing a layer of M units with $\sqrt{M}$ layers of $\sqrt{M}$ units does not change energy's quadratic scaling with unit count for deep nets laid out in 2-D.

Role of Exponents that Scale Width and Depth

I generalize this analysis to the case where a deep net's width n scales as its unit count M raised to the power β, where $0<\beta<1$. Substituting $n=(\gamma/\kappa)M^{\beta}$ in Eq. 1 yields

$$\frac{W}{Mn}=1+\frac{\kappa}{\gamma}\frac{W^{1/D}}{M^{\beta}}=1+M^{(1+\beta)/D-\beta}\left(\frac{\kappa}{\gamma}\right)^{1-1/D}\left(\frac{W}{Mn}\right)^{1/D}\iff w=1+M^{(1+\beta)/D-\beta}\left(\frac{\kappa}{\gamma}\right)^{1-1/D}w^{1/D}$$

When $(1+\beta)/D-\beta=0\Leftrightarrow\beta=1/(D-1)$, wire per synapse w is independent of the deep net's size M, confirming that scaling connections per unit n as $M^{1/(D-1)}$ keeps wire per synapse w constant. If n scales differently, then γ in Eq. 3 must be replaced by $\gamma/M^{1/(D-1)-\beta}$—or by $\gamma/M^{1-\beta}$ for $D=2$ and $\gamma/M^{1/2-\beta}$ for $D=3$.

Substituting $\gamma/M^{1-\beta}$ for γ in Eq. 3's 2-D solution yields a 2-D chip's wire per synapse w and multiplying by total number of synapses $Mn=(\gamma/\kappa)M^{1+\beta}$ yields the total wire W:

$$w = 1 + \frac{1+\sqrt{1+2\gamma/M^{1-\beta}}}{\gamma/M^{1-\beta}} \quad \text{and} \quad W = \frac{M^2}{2}\left(1+\gamma/M^{1-\beta}+\sqrt{1+2\gamma/M^{1-\beta}}\right) \tag{4}$$

For $M^{1-\beta} \gg \gamma$, $W \approx M^2/2$. This result proves that energy per inference scales as unit count squared for deep nets with arbitrary width–depth scaling laid out in 2-D.

Substituting $\gamma/M^{1/2-\beta}$ for γ in Eq. 3's 3-D solution and following the same procedure yields a 3-D chip's wire per synapse w and total wire W:

$$w = 1 + \frac{3}{2\gamma M^{\epsilon}}\left(\rho(\gamma M^{\epsilon}) + \frac{1}{\rho(\gamma M^{\epsilon})}\right) \quad \text{and} \quad W = \frac{M^{3/2}}{3^{1/2}}\left(\frac{2\gamma M^{\epsilon}}{3} + \rho(\gamma M^{\epsilon}) + \frac{1}{\rho(\gamma M^{\epsilon})}\right) \tag{5}$$

where $\epsilon = \beta - 1/2 \geq 0$ ensures $\gamma M^{\epsilon} \geq 1$; otherwise $\rho(\cdot)$ is imaginary. For $\gamma M^{\epsilon} \gg 1$, $\rho(\gamma M^{\epsilon}) \approx (2\gamma M^{\epsilon})^{1/3}$ and $W \approx 2\gamma M^{3/2+\epsilon}/3^{3/2}$. This result proves that energy per inference scales as unit count raised to at least three-halves for deep nets with arbitrary width–depth scaling laid out in 3-D.

Thermal Stability

As mentioned above, circuit-switching's energy-use scales as $W=M^2$ in 2-D and as $W=M^{3/2}$ in 3-D (see Eqs. 4 & 5), producing heat proportionately; but heat-dissipating surface area scales as $W=M^2$ in 2-D and as $W^{2/3}=M$ in 3-D. In general, heat production scales as $W \propto M^{D/(D-1)}$ whereas heat dissipation scales as $W^{2/D} \propto M^{2/(D-1)}$. Therefore, to ensure thermal stability, it is necessary for the fraction of units that signal per inference to scale as $M^{(2-D)/(D-1)}$. Multiplying this fraction by the $\sqrt{M}$ units a layer has yields $M^{(3-D)/2(D-1)}$ active units. So while all of a layer's $\sqrt{M}$ units can signal in 2-D ($D=2$), only a fixed number can do so 3-D ($D=3$). In this thermally stable case, energy per inference scales as $M^{2/(D-1)}$—or quadratically in 2-D and linearly in 3-D.

Comparison with Packet Switching

I now compare these energy and space use predictions for circuit switching in 2-D and 3-D with those for packet switching in 2-D, the current standard practice. This routing technique developed for chip multiprocessors and adapted to neuromorphic chips, shares $\log_2(n)$ wires among n units and conveys a binary-amplitude signal on each wire to communicate a unit's unary-amplitude signal[4] (Table 1). For l layers of units, $\log_2(l)$ wires are added to these $\log_2(n)$wires. Together, these wires convey $\log_2(M)$-bit addresses that identify one of $M=nl$ units[5].

By exchanging a unit's unary signal for $\log_2(M)$ binary signals, packet switching allows the l $n\times n$ crossbars to be tiled in a 2-D array and connected to their four nearest-neighbors with a grid (NSWE routing)[6,7]. A layer's n $W^{1/2}$-unit-long wires are replaced with two sets of $\log_2(n)$ n-unit-long wires that carry a row or column address along a crossbar's periphery; and four sets of $\log_2(M)$ $\{n+\log_2(n)\}$-unit-long wires that carry row, column, and layer addresses across a crossbar and its row or column address bus to four neighboring crossbars. Thus, the area projections occupy drops from $MW^{1/2}=M^2/\sqrt{2}$ square-units, substituting $W \approx M^2/2$ (see Eq. 4), to

Table 1 | Packet versus Circuit Switching

n neurons 1 spike	wires /axon	signals /spike	signals :wires
packet	$\log_2(n)/n$	$\log_2(n)$	1 : 2
circuit	1	1	1 : n

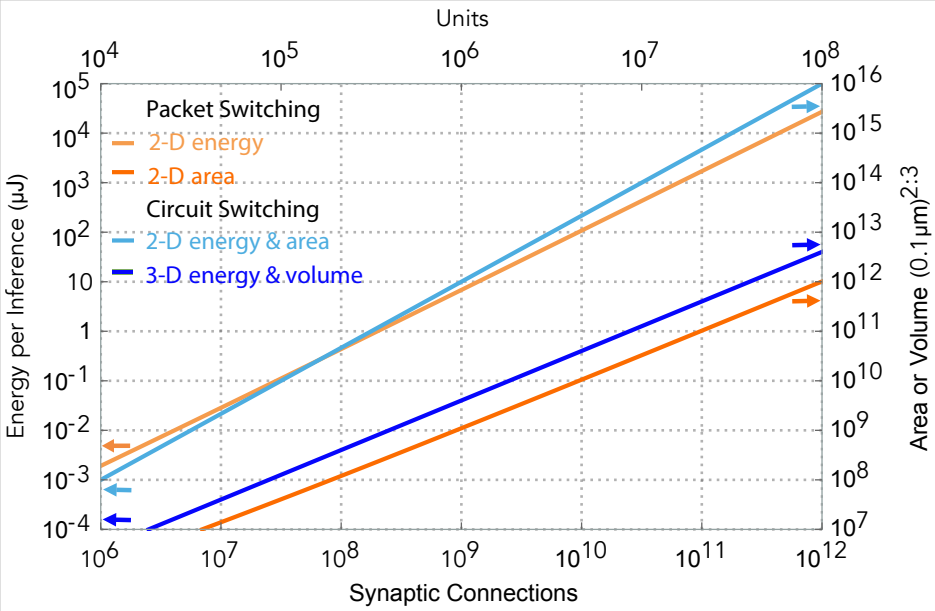

FIG. 1: How energy and space use scale for a deep net laid out in 2-D or 3-D Energy per inference (*left axis*) and 2-D area or 3-D volume (*right axis*) are plotted versus synaptic connections (*bottom axis*) and neuronal units (*top axis*). For circuit switching, a single curve describes how energy and area (*light blue*) or energy and volume (*deep blue*) scale. For packet switching, two separate curves describe how energy (*light orange*) and area (*deep orange*) scale. A wire's pitch is 0.1 μm and a signal's energy-use is 100 fJ/mm.

$$l\left(2\log_2(n)n + 4\log_2(M)(n+\log_2(n))\right) = nl\left(2\log_2(n) + 4\log_2(M) + 4\log_2(n)/n\right) \approx 5M\log_2(M)$$

square-units, substituting $nl = M$ and assuming that $2\log_2(n) \approx \log_2(M)$ and $4\log_2(n) \ll n$. This area is nearly $\sqrt{M}$-fold less than the $Mn = M^{3/2}/2$ square-units connections occupy. Thus, packet switching rescales chip area from M^2 to $M^{3/2}$ (FIG. 1, *right axis*).

When it comes to energy-use, packet switching is more efficient than circuit switching for nets with more than 2×10^5 units (FIG. 1, *left axis*). With packet switching, a projection's length scales not as M but rather as $(M^{3/2})^{1/2} = M^{3/4}$ and, for dense signaling, energy per inference scales not as M^2 but rather as $M \times M^{3/4} = M^{1.75}$. But exchanging one unary-amplitude signal for $\log_2(M)$ binary-amplitude signals increases the pre-exponential factor. As a result, only when M exceeds 2×10^5 units does packet switching becomes more energy-efficient.

However, circuit switching in 3-D is more energy-efficient and as space-efficient as packet switching in 2-D. Adding a third dimension reduces circuit-switching's total wiring (W) from $\frac{1}{2}M^2$ to $2(M/3)^{3/2}$ ($\gamma = 1$, $\epsilon = 0$; see Eqs. 4 & 5) and thus reduces its energy and space-use $\sqrt{M}$-fold. Both now scale as $M^{1.5}$—more efficient than 2-D packet-switching's energy-use ($\log_2(M)M^{1.75}$) and as efficient as 2-D packet-switching's space-use ($M^{1.5}$).

---

[2] Diorio, C., Hasler, P., Minch, A., & Mead, C. A. (1996). A single-transistor silicon synapse. *IEEE transactions on Electron Devices*, *43*(11), 1972-1980.

[3] Chklovskii, D. B. (2004). Synaptic connectivity and neuronal morphology: two sides of the same coin. *Neuron*, *43*(5), 609-617.

[4] Boahen, K. A. (2000). Point-to-point connectivity between neuromorphic chips using address events. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, *47*(5), 416-434.

[5] Benjamin, B. V. et al. (2014). Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proceedings of the IEEE*, *102*(5), 699-716.

[6] Merolla, P. A. et al. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, *345*(6197), 668-673.

[7] Davies, M. et al. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, *38*(1), 82-99.

III DENDRITE MODEL

In this section, I present differential equations used to model a spiny dendrite and parameter values used to simulate these equations. I also derive and plot this simulated model's phase portrait and energy diagram.

MODEL EQUATIONS

Two differential equations describe membrane potentials $\nu_j(t)$ and $v_j(t)$ of the j^{th} spine head and dendrite segment, respectively:

$$\tau_{\text{spi}}\dot{\nu}_j(t) = \{v_j(t) - \nu_j(t)\} + \frac{g_{\text{Kir}}}{s_{\text{Kir}}}\frac{e_{\text{K}} - \nu_j(t)}{1 + \text{e}^{-\frac{\nu_j(t) - v_{\text{iha}}}{v_{\text{isp}}}}} + \frac{g_{\text{nm}\,j}}{s_{\text{nm}}}\frac{e_{\text{cat}} - \nu_j(t)}{1 + \text{e}^{-\frac{\nu_j(t) - v_{\text{nha}}}{v_{\text{nsp}}}}} + g_{\text{am}\,j}\{e_{\text{cat}} - \nu_j(t)\}$$

$$\tau_{\text{seg}}\dot{v}_j(t) = \{\nu_j(t) - v_j(t)\} + \ell^2\{v_{j-1}(t) - 2v_j(t) + v_{j+1}(t)\} + \frac{g_{\text{Kv}}}{s_{\text{Kv}}}\frac{e_{\text{K}} - v_j(t)}{1 + \text{e}^{-\frac{v_j(t) - v_{\text{Kha}}}{v_{\text{Ksp}}}}} + \frac{g_{\text{Nav}}}{s_{\text{Nav}}}\frac{e_{\text{Na}} - v_j(t)}{1 + \text{e}^{-\frac{v_j(t) - v_{\text{Nha}}}{v_{\text{Nsp}}}}}$$

The two time-constants $\tau_{\text{spi}} = C_{\text{spi}} \big/ G_{\text{nck}}$ and $\tau_{\text{sli}} = C_{\text{seg}} \big/ G_{\text{nck}}$ are set by C_{spi}, the membrane capacitance of a spine head; C_{seg}, the membrane capacitance of a segment; and G_{nck}, the conductance of a spine neck. This conductance couples the head to the segment (first term in their equations). The length-constant $\ell = (R_{\text{seg}}G_{\text{nck}})^{-1/2}$ is set by G_{nck} and R_{seg}, the axial resistance of a segment. This resistance couples a segment to adjacent segments (second term in segment's equation). TABLE 1 gives the values for these capacitances and conductances[8,9,10].

The remaining terms model synaptic and/or active conductances. The head has an inward-rectifying potassium conductance (KIR) that hyperpolarizes it and glutamate-gated cation conductances (NMDA and AMPA) that depolarize it; the former is voltage-sensitive. The segment has a potassium conductance (KV) that hyperpolarizes it and a sodium conductance (NAV) that depolarizes it; both are voltage-sensitive. All these voltage-sensitive conductances are activated sigmoidally. TABLE 2 specifies the change in membrane potential that e-folds their activation (v_{isp}, v_{nsp}, v_{Ksp}, and v_{Nsp}), the membrane potential at which activation is equal to ½ (v_{iha}, v_{nha}, v_{Kha}, and v_{Nha}), and their conductance's value (g_{Kir}, $g_{\text{nm}\,j}$, g_{Kv}, and g_{Nav}) when the membrane potential equals the ion-channels' reversal potential (e_{K}, e_{cat}, e_{K}, and e_{Na}). This value is specified in units of G_{nck} (e.g., $g_{\text{Kir}} = G_{\text{Kir}} \big/ G_{\text{nck}}$). To accommodate that, $s_{\text{ch}} = \left(1 + \exp\left(-\frac{e_{\text{ion}} - \nu_{\text{ha}}}{\nu_{\text{sp}}}\right)\right)^{-1}$ normalizes the slope of the driving force times the sigmoidal activation to unity at the reversal potential[11].

As mentioned in the text, NMDA and KIR work together to make a spine bistable. As $\nu_j(t)$ rises from hyperpolarized levels, NMDA's conductance e-folds every 12.5 mV, attains half its maximal level when $\nu_j(t)$ crosses −23.7 mV, and reverses its current from inward to outward when $\nu_j(t)$

TABLE 1 | PASSIVE CONDUCTANCES AND CAPACITANCES

	SEGMENT	SPECIFIC VALUE	STANDARD UNITS	REMARK & REFERENCES
C_{spi}	10.4 fF	10 fF/μm²	1 μF/cm²	288-nm radius gives 0.1-μm³ volume [8]
G_{nck}	1.70 nS	250 nS/μm	400 Ω cm	1.8-μm length and 125-nm diameter [8,9]
C_{seg}	7.85 fF	10 fF/μm²	1 μF/cm²	1-μm length and 250-nm diameter [10]
G_{seg}	12.3 nS	250 nS/μm	400 Ω cm	Dimensions and reference given above

crosses 0 mV[12]. Conversely, as $\nu_j(t)$ drops from depolarized levels, KIR's conductance e-folds every −10 mV, attains half its maximal level when $\nu_j(t)$ crosses −80 mV, and reverses its current from outward to inward when $\nu_j(t)$ crosses −90 mV[12].

TABLE 2 | ACTIVE CONDUCTANCES

	v_{sp} (mV)	v_{ha} (mV)	g_{rev} (G_{nck})	e_{ion} (mV)	REF.
NMDA	12.5	−23.7	0.60	0	[12]
KIR	−10.0	−80	3.75	−90	[12]
NAV	9.0	−28.5	0.65	50	[13,18]
KV	17.0	0	0.005	−90	[14]

As mentioned in the text, NAV regenerates a plateau potential in the segment and KV limits its amplitude. As $v_j(t)$ rises from hyperpolarized levels, KV's current reverses from inward to out-ward when $v_j(t)$ crosses −90 mV, KV's and NAV's conductances e-fold every 17 and 9 mV, respectively, attain half their maximal levels when $v_j(t)$ crosses 0 and −28.5 mV, and NAV's current reverses from inward to outward when $v_j(t)$ crosses 50 mV[13,14]. Increasing NAV's specified conductance, g_{Nav}, lowers the threshold $v_j(t)$ must exceed to initiate a plateau potential. If g_{Nav} is too large, however, NAV will sustain this plateau—on its own—after NMDA's conductance, $g_{nmj}(t)$, decays.

PARAMETER VALUE CHOICES

I chose these channels' conductances based on recent experimental findings and models thereof:

NMDA: Spiny synapses are surprisingly strong (1–2.5 nS), with an AMPA-to-neck conductance -ratio of 0.86 (median $G_{syn} = 1.47$ nS; mean $G_{nck} = 1.7$ nS)[9]. My model's NMDA conductance, $g_{nmj} = 0.60 G_{nck} = 1.02$ nS, corresponds to a NMDA:AMPA ratio of 1:1.47, which is typical[15].

KIR: Cortical pyramidal cells do receive inhibition on their dendritic spines ($GABA_B$/KIR)[16,17]. In the model, this placement enhanced bistability without dampening the amplifying action of the segment's NAV channels, thanks to the spine-neck's resistance. The strength of these inhibitory synapses has not been measured.

NAV: Previous modeling replicated NMDA plateau potentials and Na^+ spikelets in thin dendrites—measured using voltage-sensitive dyes—with 375 pS/μm² of voltage-gated sodium channels[18]. My model has almost four times that: $g_{Nav} = 0.65 G_{nck}$ corresponds to 1.41 nS/sq-μm.

KV: Apical tuft sites express voltage-gated potassium channels at a density of 23 pS/sq-μm (sustained component)[14]. The model has about half that: $g_{Kv} = 0.005 G_{nck}$ corresponds to 10.8 pS/sq-μm.

Although my model requires higher sodium channel density than previous models, the value it predicts for highly selective sequence discrimination (1.41 nS/μm²) is well within the physiological range. For reference, the sodium channel density at an axon-hillock is over an order-of-magnitude higher (30 nS/μm²)[13].

SIMULATIONS

I found that spikes delivered consecutively to spines spaced 3 μm apart advance a plateau potential along the modeled dendrite. To obtain this result, I lumped three 1-μm segments together, replaced membrane potentials of these lumped segments and of their spine heads with averages—$v_k(t) = \langle v_j(t) \rangle$ and $\nu_k(t) = \langle \nu_j(t) \rangle$ for $sk \le j < s(k+1)$ for lumps of s segments—and all conductances with averages. The same equations above describe these lumped segments, except that ℓ / s replaces ℓ, with $s = 3$.

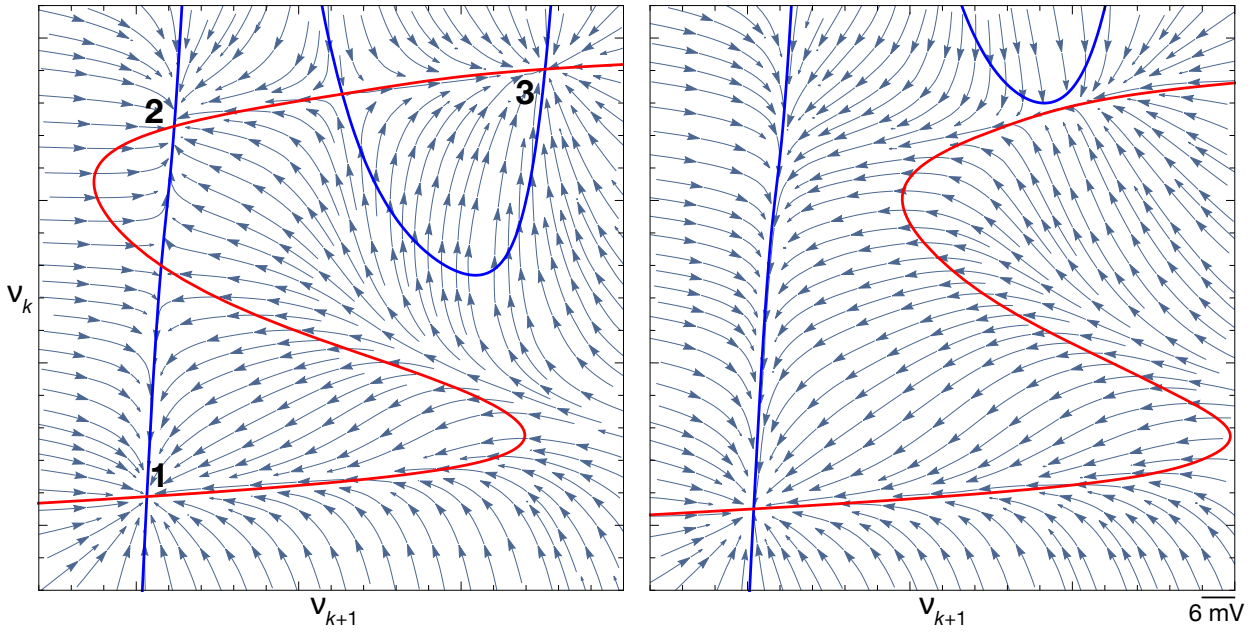

FIG. 2 | Phase portraits of membrane potentials of spines k and $k+1$
A vector $(\dot{v}_{k+1}, \dot{v}_k)$ at coordinates (v_{k+1}, v_k) plots the membrane potentials' rate of change at those values. A line plots the membrane potentials' values where one is stationary ($\dot{v}_k = 0$ in *red* and $\dot{v}_{k+1} = 0$ in *blue*). These "nullclines" intersect where both potentials are stationary—a stable, metastable (saddle), or unstable point.
Left – Peak NMDA conductances ($g_{\mathrm{nm}k} = g_{\mathrm{nm}k+1} = 0.6$): There are three stable points (v_k-v_{k+1}): (1) rest-rest (−88 mV, − 85 mV), (2) plateau-rest (−83 mV, − 18 mV), and (3) plateau-plateau (−14 mV, − 6.0 mV). And also two saddle points: One separates rest-rest from plateau-rest and the other separates plateau-rest from plateau-plateau. All trajectories converge to a stable point, but some trajectories initially converge to a saddle point and then diverge.
Right – Reduced NMDA conductances ($g_{\mathrm{nm}k} = g_{\mathrm{nm}k+1} = 0.09$): The nullclines become less *N*-shaped and, as a result, plateau-rest disappears and then plateau-plateau disappears, both through "saddle-node bifurcations". This sequence explains the cascade of plateau-to-rest transitions in FIG 3c of the main text.

I strung together twenty of these 3-µm segments to examine the spike-sequence selectivity of a short stretch of dendrite. Spikes arrive at 7-ms intervals and each one activates a spine's AMPA and NMDA conductances. To guarantee the first segment—at the dendrite tip—transitions from rest to plateau, its KIR conductance was omitted. I solved the differential equations of this model numerically for time-varying synaptic conductances using MATHEMATICA 12.0.

Time-varying AMPA and NMDA conductances were modeled as follows. $g_{\mathrm{am}k}(t)$ varies as $(t/\tau_0^2)\mathrm{e}^{-\frac{t}{\tau_0}}$, rising and decaying with time constant $\tau_0 = 3$ ms. $g_{\mathrm{nm}k}(t)$ varies as $0.5\mathrm{e}^{-\frac{t}{\tau_2}} + \mathrm{e}^{-\frac{t}{\tau_3}} + \mathrm{e}^{-\frac{t}{\tau_4}} - 2.5\mathrm{e}^{-\frac{t}{\tau_1}}$, rising with time constant $\tau_1 = 1.4$ ms and decaying with time constants $\tau_2 = 12.6$, $\tau_3 = 57.3$, and $\tau_4 = 372$ ms (after adjusting from 23 to 35°C with $q_{10} = 3$)[19]. The most rapidly decaying component's amplitude was set to half that of the other two, which were equal in amplitude. These waveforms were scaled to peak at $g_{\mathrm{nm}k} = 0.60$ and $g_{\mathrm{am}k} = 0.25$. All membrane voltages were initialized to their values at rest.

PHASE PORTRAIT AND ENERGY DIAGRAM

My analysis focused on two adjacent segments. Their distal and proximal neighbors were replaced with the (model) dendrite's equivalent conductance, tied to −1.5 mV (plateau) and −90 mV (rest), respectively. I reduced this four-dimensional model to a two-dimensional one by

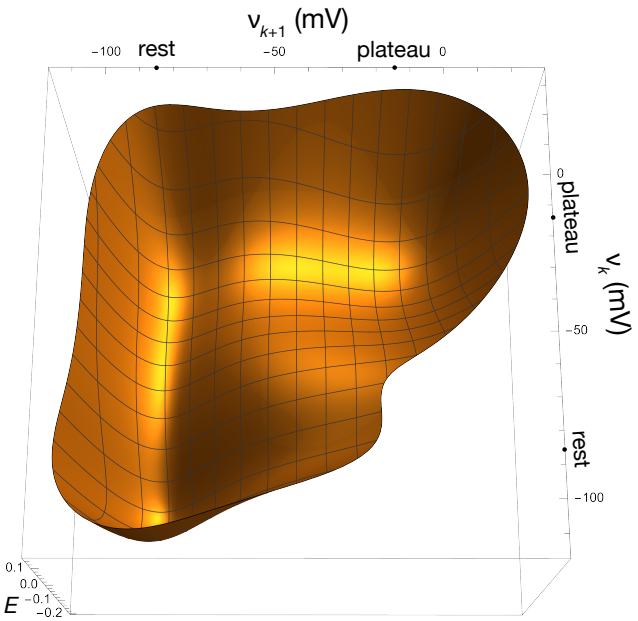

FIG. 3 | Energy landscape of spines k and $k+1$ Electrical energy E versus voltages ν_k and ν_{k+1} of spines k and $k+1$. When spine $k+1$ is at rest, increasing ν_k minimizes $E(\nu_k, \nu_{k+1})$ at two points. These equilibria correspond to the rest-rest and plateau-rest stable points (ν_k-ν_{k+1}); a medium energy barrier separates them. When spine k is at plateau, increasing ν_{k+1} also minimizes $E(\nu_k, \nu_{k+1})$ at two points. These equilibria correspond to the plateau-rest and plateau-plateau stable points; a small energy barrier separates them. Increasing ν_k and ν_{k+1} together also minimizes $E(\nu_k, \nu_{k+1})$ at two points. These equilibria correspond to the rest-rest and plateau-plateau stable points; a large energy barrier separates them. It is fourfold higher than that from plateau-rest to plateau-plateau. Thus, transitioning spine k from rest to plateau lowers the energy barrier for transitioning spine $k+1$ in that direction fourfold.

solving for the dendrite segments' voltages at equilibrium ($\dot{\upsilon}_k = 0$ and $\dot{\upsilon}_{k+1} = 0$), as their time-constant ($\tau_{\text{seg}} = 4.6$ μs) is orders of magnitude shorter than the timescale on which synaptic inputs vary. A phase portrait of this 2-D dynamical system reveals that it has three stable points (FIG 2).

To obtain the 2-D system's energy function $E(\nu_k, \nu_{k+1})$, I first obtained one for the full 4-D system $E(\nu_k, \upsilon_k, \nu_{k+1}, \upsilon_{k+1})$. This function's derivative with respect to any variable gives that variable's temporal derivative (e.g., $\dot{\nu}_k = -dE/d\nu_k$)[20]. Substituting υ_k's and υ_{k+1}'s equilibrium solutions reduces $E(\nu_k, \upsilon_k, \nu_{k+1}, \upsilon_{k+1})$ to $E(\nu_k, \nu_{k+1})$. This 2-D energy landscape has three wells that correspond to the three stable points (FIG 3).

---

[8] Kasthuri, N., et al. (2015). Saturated reconstruction of a volume of neocortex. *Cell*, *162*(3), 648-661.

[9] Beaulieu-Laroche, L., & Harnett, M. T. (2018). Dendritic spines prevent synaptic voltage clamp. *Neuron*, *97*(1), 75-82.

[10] Liao, M., Liang, X., & Howard, J. (2021). The narrowing of dendrite branches across nodes follows a well-defined scaling law. *Proceedings of the National Academy of Sciences*, *118*(27).

[11] Take the derivative with respect to $v_j(t)$, substitute e_{ion} for $v_j(t)$, and set s_{ch} equal to the result.

[12] Sanders, H., Berends, M., Major, G., Goldman, M. S., & Lisman, J. E. (2013). NMDA and GABAB (KIR) conductances: the "perfect couple" for bistability. *Journal of Neuroscience*, *33*(2), 424-429.

[13] Mainen, Z. F., & Sejnowski, T. J. (1996). Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature*, *382*(6589), 363-366.

[14] Harnett, M. T., Xu, N. L., Magee, J. C., & Williams, S. R. (2013). Potassium channels control the interaction between active dendritic integration compartments in layer 5 cortical pyramidal neurons. *Neuron*, *79*(3), 516-529.

[15] Eyal, G., et al. (2018). Human cortical pyramidal neurons: From spines to spikes via models. *Frontiers in Cellular Neuroscience*, *12*, 181.

[16] Koyrakh, L., et al. (2005). Molecular and cellular diversity of neuronal G-protein-gated potassium channels. *Journal of Neuroscience*, *25*(49), 11468-11478.

[17] Kulik, Á., et al. (2006). Compartment-dependent colocalization of Kir3.2-containing K+ channels and GABAB receptors in hippocampal pyramidal cells. *Journal of Neuroscience*, *26*(16), 4289-4297.

[18] Gao, P. P., et al. (2021). Local glutamate-mediated dendritic plateau potentials change the state of the cortical pyramidal neuron. *Journal of Neurophysiology*, *125*(12), 23-42.

[19] Iacobucci, G. J., & Popescu, G. K. (2017). NMDA receptors: linking physiological output to biophysical operation. *Nature Reviews Neuroscience*, *18*(4), 236-249.

[20] Tank, D. W., & Hopfield, J. J. (1987). Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences*, *84*(7), 1896-1900.