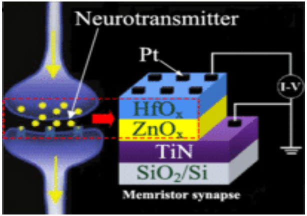
Memristor drift-diffusion systems for brain-inspired neuromorphic computing

Ansgar Jüngel
Technische Universität Wien, Austria
https://www.tuwien.at/mg/asc/juengel

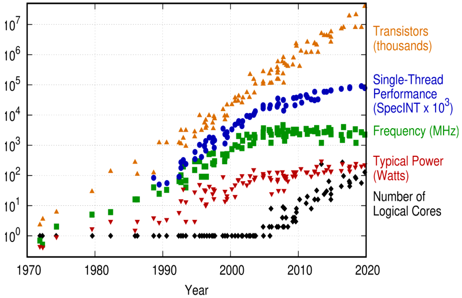
European Research Council
Established by the European Commission
Der Wissenschaftsfonds.



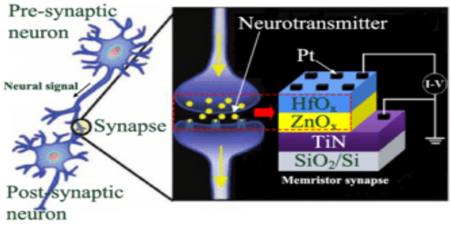
Analogy between synapse and memristor device

Semiconductor devices

- Basic element in computer: semiconductor transistor
- 3 nm technology (2022): 48 nm contacted gate pitch → physical limits, leakage current
- Novel technologies needed!
- Neuromorphic computing: mimic neurobiological networks
- New concept by Mead (1990): memristor = nonlinear resistor with memory
- Pros: Ultra-low power consumption, nanosize, nonvolatile memory

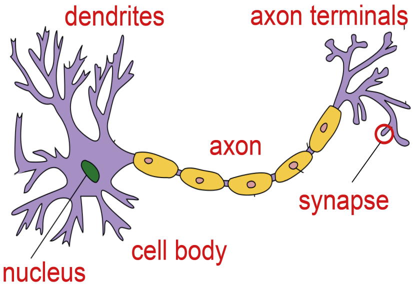


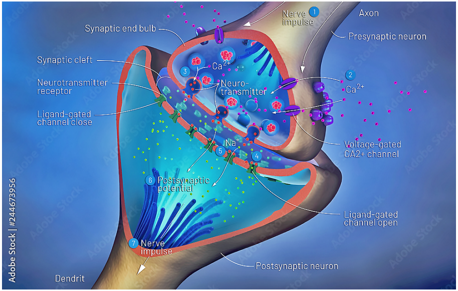
Transistor number in microchips ©Karl Rupp

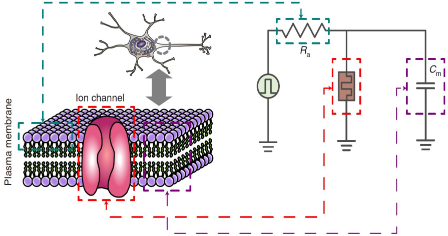


©Wang et al., Nanoscale Res. Lett. 2017

Synapse – ion channel – electric circuit







Axon network:

→ synapses → ion channel

→ memristors → electric circuits

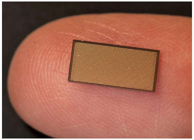
→ novel brain-inspired algorithms

→ neuromorphic computing

Neuromorphic computing

Neuromorphic chip = inspired by biological neural network in brain

- 2014 IBM TrueNorth: 1 million artificial neurons, 256 millions synapses
- 2017 Intel Loihi 1: 130,000 artificial neurons, 130 million synapses



©Intel Loihi 2 at Wikipedia

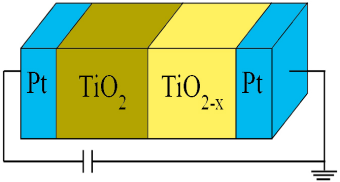
- 2021 Intel Loihi 2: 1 million artificial neurons, 120 millions synapses
- 2022 Stanford NeuRRAM: 3 million RRAM cells
- 2024 Hala Point: 1152 Loihi 2 processors, 1.15 billion neurons, 128 billion synapses, total 2600 Watt power consumption

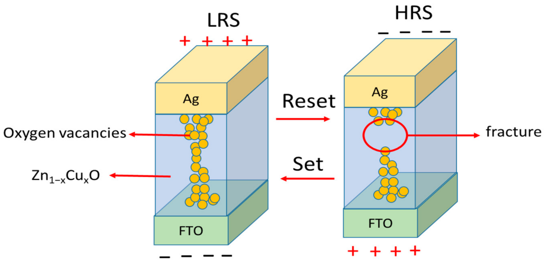
→ human brain: 86 billion neurons, 100 trillions of synapses, 25 Watt

Pros: Very low power 0.1 W, ultrasmall size, merges memory & computing

Cons: High production precision needed, incompatible with standard programming and data formats

Oxide-based memristor device





©Weng et al., nanomaterials 2023

- Oxide-based memristor = titanium dioxide film sandwiched between two metal electrodes (Strukov et al. 2008)
- TiO_{2-x} = Ti oxide doped with oxide vacancies (loss of oxygen atoms)
- External electric field changes length of doped TiO_{2-x} region
- Open state: doped region has maximal length → minimal resistance
- Closed state: doped region has minimal length → maximal resistance

Overview

Memristor drift-diffusion equations

$$\partial_t n = \operatorname{div} J_n, \quad J_n = \nabla n - n\nabla V$$
$$\partial_t p = -\operatorname{div} J_p, \quad J_p = -(\nabla p + p\nabla V)$$
$$\partial_t D = -\operatorname{div} J_D, \quad J_D = -(\nabla D + D\nabla V)$$
$$\lambda^2 \Delta V = n - p - D + A(x) \quad \text{in } \Omega \subset \mathbb{R}^d, \ t > 0$$

- Densities for electrons n, holes p, oxide vacancy D; electr. potential V
- Mixed boundary conditions: $\bar{n}, \bar{p}, \bar{V}$ on $\Gamma_{\rm Dir}$, no flux on $\Gamma_{\rm Neu}$
- No-flux boundary conditions for D on $\partial\Omega$

Mathematical difficulties:

- Non-fitting boundary conditions: Issues when integrating by parts
- Three species: Multiply eqs. by n and p, assume $\overline{n} = \overline{p} = 0$, fixed D

$$\frac{1}{2}\frac{d}{dt}\int_\Omega (n^2 + p^2)dx + \int_\Omega (|\nabla n|^2 + |\nabla p|^2)dx = \int_\Omega (n\nabla n - p\nabla p)\cdot\nabla V dx$$
$$= -\frac{1}{2\lambda^2}\int_\Omega \underbrace{(n^2 - p^2)(n-p}_{\geq 0} - D + A)dx \leq C(A,D)\int_\Omega (n^2+p^2)dx$$

Drift-diffusion equations: state of the art

Two species:

$$\begin{aligned} \partial_t n &= \operatorname{div} J_n, \quad J_n = \nabla n - n\nabla V \\ \partial_t p &= -\operatorname{div} J_p, \quad J_p = -(\nabla p + p\nabla V) \\ \lambda^2 \Delta V &= n - p - D(x) + A(x) \quad \text{in } \Omega \subset \mathbb{R}^3,\ t > 0 \end{aligned}$$

- Fick's diffusion law for charged particles: Nernst-Planck equations
- Semicond.: Van Roosbroeck 1950, Mock 1974, Gajewski-Gröger 1986

More than two species:

- Four-species system (drift only for n, p): Verri-Porro-Sacco-Salsa '18 $\rightarrow$ use monotonicity argument
- Global existence of solutions in $\mathbb{R}^2$: Bothe-Fischer-Saal 2014, Glitzky-Hünlich 1997/2005
- Global existence in $\mathbb{R}^3$: Bothe-Fischer-Pierre-Rolland 2014 (no-flux/Robin boundary conditions)

Why are the restrictions necessary?

- Idea of Glitzky-Hünlich 1997: Hölder ($2 < q \le q_0$, $q' = 2q/(q-2)$) & generalized Gagliardo-Nirenberg's inequality

$$\begin{aligned}\int_\Omega n\nabla n \cdot \nabla V dx &\le \|n\|_{L^{q'}(\Omega)}\|\nabla n\|_{L^2(\Omega)}\|\nabla V\|_{L^q(\Omega)} \\ &\le C\|n\|^{\theta}_{H^1(\Omega)}\|n\|^{1-\theta}_{L^1(\Omega)} \cdot \|\nabla n\|_{L^2(\Omega)} \cdot \big(1 + \|n-p\|_{L^{2q/(q+2)}(\Omega)}\big) \\ &\le \delta\|\nabla n\|^{1+(3d-2)/(d+2)}_{L^2(\Omega)} + C\big(\delta, n, p\big), \quad \delta > 0\end{aligned}$$

  where C depends on $L^1 \log L^1$ norm of n, p; $1 + \frac{3d-2}{d+2} \le 2 \Leftrightarrow d \le 2$

- Idea of Bothe et al. 2014: Use free energy and estimate

$$\begin{aligned}\int_\Omega \nabla(n-p-D) \cdot \nabla V dx &= -\int_\Omega \underbrace{(n-p-D)}_{=\lambda^2\Delta V}\Delta V dx \\ &\quad + \int_{\partial\Omega}(n-p-D)\nabla V \cdot \nu dx\end{aligned}$$

  bounded if $\nabla V \cdot \nu$ can be controlled on $\partial\Omega$ $\to$ no mixed b.c.!

Third idea: entropy method

$$\partial_t n = \mathrm{div}(\nabla n - n\nabla V), \quad \partial_t p = \mathrm{div}(\nabla p + p\nabla V)$$
$$\partial_t D = \mathrm{div}(\nabla D + D\nabla V), \quad \lambda^2 \Delta V = n - p - D + A(x)$$

Objective: Allow for any space dim. & mixed boundary conditions $(\bar{n}, \bar{p}, \bar{V})$

- Free energy = sum of thermodynamic entropies and electric energy:

$$H = \int_\Omega \left\{ n\left(\log\frac{n}{\overline{n}} - 1\right) + p\left(\log\frac{p}{\overline{p}} - 1\right) + D(\log D - 1 + \overline{V})\right\} dx + \frac{\lambda^2}{2}\int_\Omega |\nabla(V - \overline{V})|^2 dx$$

- Free energy dissipation inequality:

$$\frac{dH}{dt} + \int_\Omega \left(n|\nabla(\log n - V)|^2 + p|\nabla(\log p + V)|^2 + D|\nabla(\log D + V)|^2\right) dx \le C(\overline{n}, \overline{p}, \overline{V})$$

- A priori estimates for n, p, D in $L^1 \log L^1(\Omega)$ and for V in $H^1(\Omega)$

$$|\nabla\sqrt{n}| = \frac{1}{2}\sqrt{n}|\nabla \log n| \le C\sqrt{n}|\nabla(\log n - V)| + C\sqrt{n}|\nabla V| \in L^1(\Omega)$$

- Implies estimates for $\sqrt{n}$, $\sqrt{p}$, $\sqrt{D}$ in $W^{1,1}(\Omega)$ only

Global existence of weak solutions

$$\partial_t n = \operatorname{div}(\nabla n - n\nabla V), \quad \partial_t p = \operatorname{div}(\nabla p + p\nabla V)$$
$$\partial_t D = \operatorname{div}(\nabla D + D\nabla V), \quad \lambda^2 \Delta V = n - p - D + A(x)$$

Theorem (Jourdana-AJ-Zamponi 2022)

Let $T > 0$, $\Omega \subset \mathbb{R}^d$ *with* $d \geq 1$, $\partial\Omega$ *Lipschitz.* $\Rightarrow \exists$ *solution* (n, p, D, V)

$$n,\ p,\ D \in L^\infty(0, T; L^1 \log L^1(\Omega)),\ \sqrt{n},\ \sqrt{p},\ \sqrt{D} \in L^2(0, T; W^{1,1}(\Omega)),$$
$$J_n,\ J_p,\ J_D \in L^1(0, T; L^1(\Omega)), \quad V \in L^\infty(0, T; H^1(\Omega))$$

If $d \leq 2$ *then* $n, p, D \in L^\infty(0, \infty; L^\infty(\Omega))$.

- Problem: $\sqrt{n_k} \in W^{1,1}(\Omega)$ not sufficient for compactness
- Prove $n_k \in W^{1,r}_{\mathrm{loc}}(\Omega)$ for $r > 1$ & apply Aubin-Lions compactness $\Rightarrow$ subsequence of (n_k) converges strongly in $L^s_{\mathrm{loc}}(\Omega)$ for $1 < s < r$
- Theorem of de la Vallée-Poussin: (n_k) converges weakly in $L^1(\Omega)$
- Cantor diagonal argument: (n_k) converges strongly in $L^1(\Omega)$
- How to prove L^∞ bounds for $d = 3$? Solution: degenerate equations

Overview

Degenerate drift-diffusion equations

$$\partial_t n = \operatorname{div}(\nabla n^{\alpha_n} - n\nabla V), \quad \partial_t p = \operatorname{div}(\nabla p^{\alpha_p} + p\nabla V)$$
$$\partial_t D = \operatorname{div}(\nabla D^{\alpha_D} + D\nabla V), \quad \lambda^2 \Delta V = n - p - D + A(x)$$

- Mixed Dirichlet–Neumann conditions $(\bar{n}, \bar{p}, \bar{V})$, no-flux cond. for D
- Electrons satisfy Fermi–Dirac statistics: $n = F_{1/2}(\mu)$
- Diffusion flux $J = n\nabla\mu$ with $n = \mu^{3/2}$ (high density regime) $\Rightarrow$ $J = n\nabla n^{2/3} = \frac{2}{5}\nabla n^{5/3}$
- Motivates α_n, $\alpha_p > 1$; gain of regularity requires $\alpha_D > 1$!

Enthalpies and free energy:

$$h_n(n) = \frac{n(n^{\alpha_n - 1} - \bar{n}^{\alpha_n - 1})}{\alpha_n - 1}, \quad h_p(p) = \frac{p(p^{\alpha_p - 1} - \bar{p}^{\alpha_p - 1})}{\alpha_p - 1}, \quad h_D(D) = \frac{D^{\alpha_D - 1}}{\alpha_D - 1}$$
$$H = \int_\Omega \left(h_n(n) + h_p(p) + h_D(D) + \frac{\lambda^2}{2}|\nabla(V - \overline{V})|^2 \right) dx$$

Free energy estimate

$$\partial_t n = \mathrm{div}(\nabla n^{\alpha_n} - n\nabla V), \quad \partial_t p = \mathrm{div}(\nabla p^{\alpha_p} + p\nabla V)$$
$$\partial_t D = \mathrm{div}(\nabla D^{\alpha_D} + D\nabla V), \quad \lambda^2\Delta V = n - p - D + A(x)$$

- Free energy:

$$H = \int_\Omega \left(h_n(n) + h_p(p) + h_D(D) + \frac{\lambda^2}{2}|\nabla(V - \overline{V})|^2 \right) dx$$

- Free energy dissipation inequality: $c_j = \alpha_j/(\alpha_j - 1)$

$$\frac{dH}{dt} + \int_\Omega \left(n|\nabla(c_n n^{\alpha_n - 1} - V)|^2 + p|\nabla(c_p p^{\alpha_p - 1} + V)|^2 + D|\nabla(c_D D^{\alpha_D - 1} + V)|^2 \right) dx$$
$$+ \frac{\lambda}{2}\int_\Omega |\nabla(V - \bar{V})|^2 dx \le C(\overline{n}, \overline{p}, \overline{V})$$

- A priori estimates for $n, p, D \in L^1 \log L^1(\Omega)$ and $V \in H^1(\Omega)$

$$\|\nabla n^{\alpha_n - 1/2}\|_{L^2} \le C \underbrace{\|\sqrt{n}(\nabla(c_n n^{\alpha_n} - V))\|_{L^2}}_{\text{uniformly bounded}} + \underbrace{\|\sqrt{n}\|_{L^6}}_{???} \underbrace{\|\nabla V\|_{L^3}}_{???}$$

Uniform bounds

$$\|\nabla n^{\alpha_n-1/2}\|_{L^2} \le C\big\|\sqrt{n}(\nabla(c_n n^{\alpha_n} - V))\|_{L^2} + \|\sqrt{n}\|_{L^6}\|\nabla V\|_{L^3}$$

- Use Gagliardo-Nirenberg inequality:

$$\|\sqrt{n}\|_{L^6} \le C\|\nabla n^{\alpha_n-1/2}\|_{L^2}^{\theta_1}\|n\|_{L^{\alpha_n}}^{\theta_2} + C\|n\|_{L^{\alpha_n}}$$

$$\|\nabla V\|_{L^3} \le C\|n\|_{L^{3/2}} + C \le C\|\nabla n^{\alpha_n-1/2}\|_{L^2}^{\theta_3}\|n\|_{L^{\alpha_n}}^{\theta_4} + C\|n\|_{L^{\alpha_n}}$$

- Insert in gradient estimate: $\theta_1 + \theta_3 < 1$ iff $\alpha_n > 6/5$

$$\|\nabla n^{\alpha_n-1/2}\|_{L^2} \le C\|\nabla n^{\alpha_n-1/2}\|_{L^2}^{\theta_1+\theta_3} + C$$

- Yields uniform bounds for ∇n and $\partial_t n$ if $\alpha_n \le 2$

Assumptions: $\frac{6}{5} < \alpha_n \le 2$ and $\|\nabla V\|_{L^3} \le C\|n\|_{L^{3/2}} + C$ (regularity)

Elliptic regularity: $\Delta V = f$ in Ω, $V = \bar{V}$ on Γ_{Dir}, $\nabla V \cdot \nu = 0$ on Γ_{Neu}

- Gröger 1994: $V \in W^{1,r}(\Omega)$ for some $r > 2$ in Lipschitz domains
- Shamir 1968: counterexample for $r \ge 4$ with smooth data
- Disser–Rehberg 2015: angle between Γ_{Dir} and $\Gamma_{\mathrm{Neu}} \le \pi \Rightarrow r > 3$

Existence of solutions

$$\partial_t n = J_n, \quad J_n = \nabla n^{\alpha_n} - n\nabla V \quad \text{and equations for } p, D, V$$

Theorem (Jüngel-Vetter 2023)

Let $d = 3$, $\frac{6}{5} < \alpha_n \le 2$ *and* $V \in W^{1,3}(\Omega)$. *Then* $\exists$ *solution* (n, p, D, V)

$$n^{\alpha_n}, p^{\alpha_p}, D^{\alpha_D} \in L^\infty(0, T; L^1(\Omega))$$

$$n^{\alpha_n - 1/2}, p^{\alpha_p - 1/2}, D^{\alpha_D - 1/2} \in L^2(0, T; H^1(\Omega))$$

$$J_n, J_p, J_D \in L^2(0, T; L^1(\Omega)), \quad V \in L^\infty(0, T; H^1(\Omega))$$

Theorem (Jüngel-Vetter 2023)

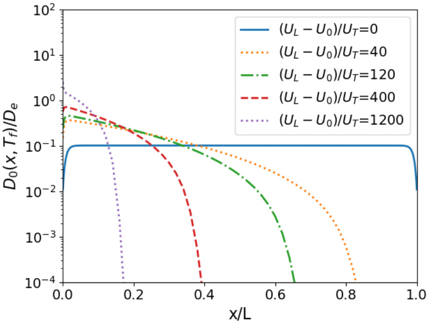
- $\alpha_n > \frac{11+\sqrt{37}}{14} \approx 1.22$, $V \in W^{1,3}(\Omega)$: $n \in L^\infty(0, T; L^q(\Omega))$ $\forall q < \infty$
- $\alpha_n > \frac{11+\sqrt{37}}{14} \approx 1.22$, $V \in W^{1,r}(\Omega)$, $r > 3$: $n \in L^\infty(0, T; L^\infty(\Omega))$

Proof: bootstrap $n \in L^1(\Omega) \to n \in L^{3/2}(\Omega)$ via linear difference equation $\to n \in L^\gamma(\Omega)$ uniformly in $\gamma \to n \in L^\infty(\Omega)$

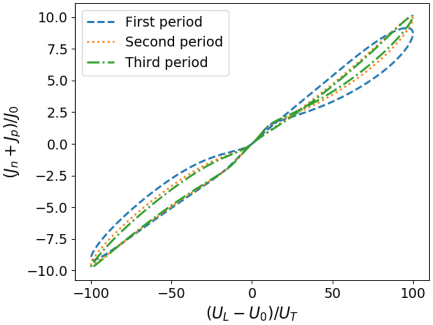
Overview

1. Introduction
2. Drift-diffusion equations
3. Degenerate drift-diffusion equations
4. Numerical simulation

Numerical simulation



Vacancy density D_0 $T = 100\,\text{ps}$



Current-voltage curve

- Reduced model, Scharfetter-Gummel scheme, finite volumes, Newton
- Left: Applied potential causes mobile vacancies to drift, complete vacancy depletion at $x = L \to$ consistent with physical expectation
- Right: Sinusoidal applied voltage $U_L(t)$, loop is fingerprint for memristor, indicates memory effect

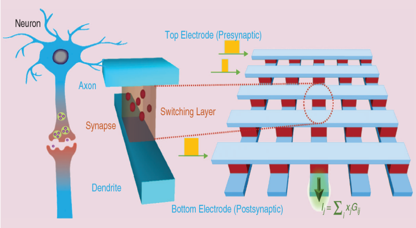
Summary and perspectives

Summary:

- Memristor drift-diffusion eqs. represents synapse in neural network
- Global existence of weak solutions in any space dimension $d \geq 1$
- Boundedness for $d \leq 2$ (linear diff.) and $d = 3$ (degenerate diff.)
- Key ideas: free energy, local/global compactness, iteration arguments
- Numerical experiments for 1D memristors show hysteresis loop

Perspectives:

- Structure-preserving numerical schemes in two dimensions
- Couple memristor model with circuit equations
- Neural plasticity & self-learning in memristor circuits



©Jeong–Lu, IEEE Nanotec. Magazine 2018

Neuromorphic computing yields novel hardware approach for AI