# Maximal Entropy Random Walk

the most random of random walks
(maximizing entropy production)

RW for minimal information about a system
in agreement with the maximal uncertainty principle.
**strong localization property**, **scale-free, nonlocal**

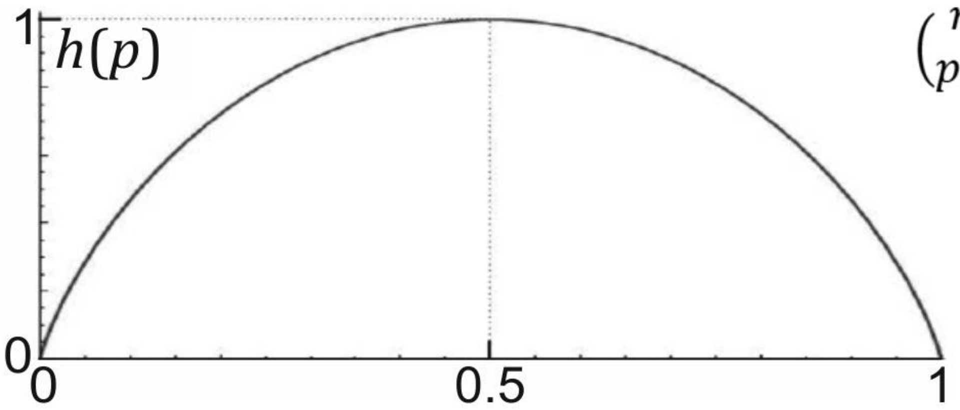
Some applications:

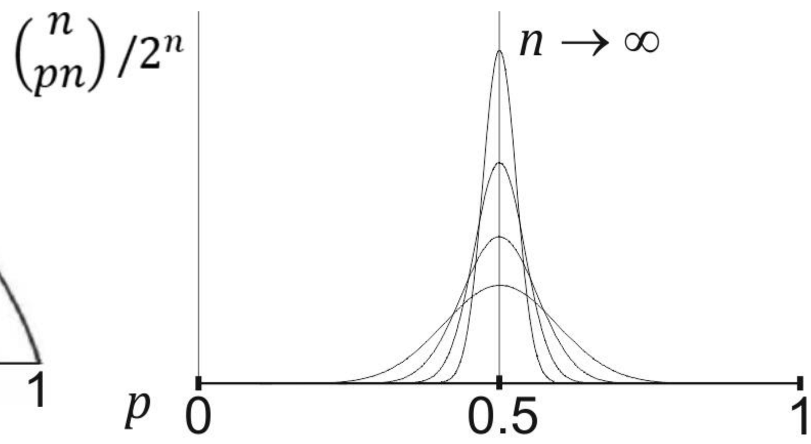
- **maximizing informational capacity** of channel under some constraints (data storage/transmission, maybe linguistics (?)),
- corrections to **diffusion models** to get agreement with quantum predictions (diffusion, conductance, molecular dynamics),
- **metrics for complex networks** (e.g. centrality measure, saliency regions, PageRank, SimRank, community detection)

Jarosław Duda, Kraków, 24.11.2014

# We need $n$ bits of information to choose one of $2^n$ possibilities.

For length $n$ 0/1 sequences with $pn$ of "1", how many bits we need to choose one?




$$\binom{n}{pn} \approx 2^{nh(p)} \qquad \text{for} \qquad h(p) := -p\lg(p) - (1-p)\lg(1-p)$$

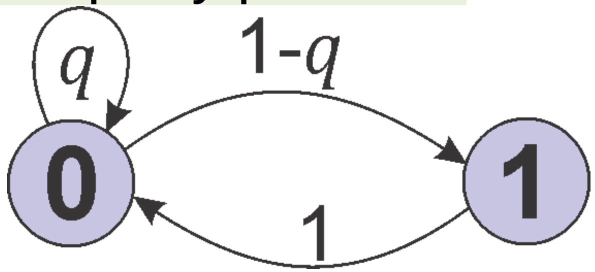
A sequence of symbols with $(p_s)_{s=0..m-1}$ probability distribution contains asymptotically $\boldsymbol{H = \sum_s p_s \lg(1/p_s)}$ **bits/symbol** ($H \le \lg(m)$)

Seen as weighted average:

**symbol/event of probability $p$ contains $\lg(1/p)$ bits.**

Fibonacci coding – as a bit sequence with **constraints**: no two neighboring '1's
e.g. 0010101000010101001001 – each sequence should be equally probable
What about statistics of a single step?

$$M = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \qquad S = \begin{pmatrix} q & 1-q \\ 1 & 0 \end{pmatrix} \qquad q = ?$$

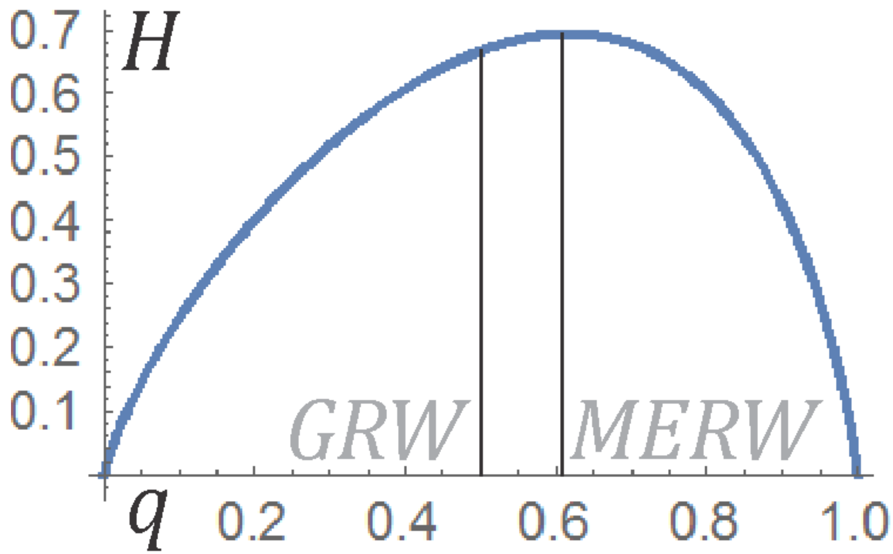



**What $q$ should we choose to maximize informational capacity?**

Stationary probability: $\pi = (\Pr(0), \Pr(1))^T$

$$\pi S = \pi$$

$$\pi = \left(\frac{1}{2-q}, 1 - \frac{1}{2-q}\right)$$




**Entropy – informational content:**

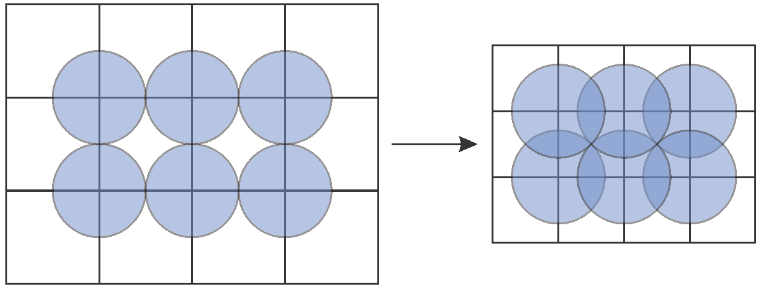
$$H = \sum_i \pi_i \sum_j S_{ij} \lg(1/S_{ij}) = \pi_0 \cdot h(q)$$

$$H_{max} \approx 0.694241913 \text{ bits/node}$$

for $q = \frac{(\sqrt{5}-1)}{2} \approx 0.618034$

My original MERW motivation: **maximizing capacity under constraints**
for 2D analogue of Fibonacci coding (“hard square”: no two neighboring ‘1’s)
We get $\boldsymbol{H} \approx \mathbf{0.58789}$ bits/node

Some application:
use magnetic dots (twice) more densely,
at cost of constraints – two dots cannot overlap.
$2 \cdot 0.58789 \approx 1.\mathbf{176}$



**We get 17.6% capacity increase due to better positioning!**
(e.g. using 1D MERW on the space of possible succeeding lines)

We need to find MERW for general situation:

| **Graph** $(M)$ | **stochastic matrix** $(S)$ | **stationary probability** $(\pi)$ |
|---|---|---|
| $M_{ab} \in \{0,1\}$ | $0 \le S_{ab} \le M_{ab}$ , $\forall_a \sum_b S_{ab} = 1$ $\rightarrow$ | $\sum_a \pi_a S_{ab} = \pi_b$ |

Average **entropy** production per step: $\sum_a \pi_a \sum_b S_{ab} \lg(1/S_{ab})$

**What $S$ should we choose? Such that each path/code is equally probable!**

Can language be seen this way –
as maximizing channel capacity under some constraints (redundancy)?

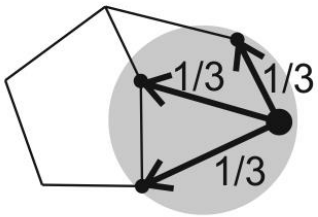
**Graph** $(M)$  **stochastic matrix** $(S)$ $\rightarrow$ **stationary probability** $(\pi)$

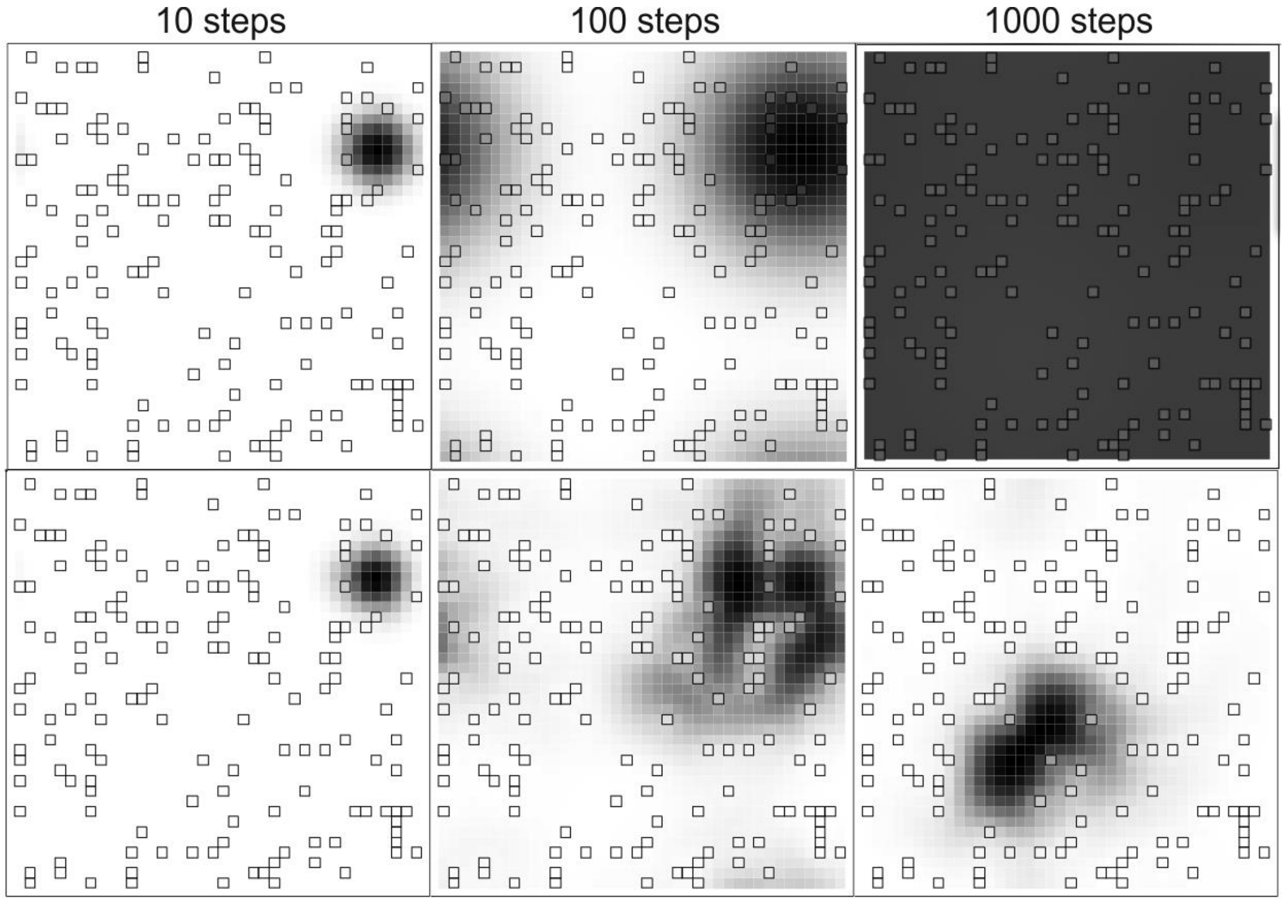
$M_{ab} \in \{0,1\}$ $\quad 0 \le S_{ab} \le M_{ab}\ ,\ \forall_a \sum_b S_{ab} = 1 \quad \sum_a \pi_a S_{ab} = \pi_b$

Average **entropy** production per step: $\sum_a \pi_a \sum_b S_{ab} \lg(1/S_{ab})$

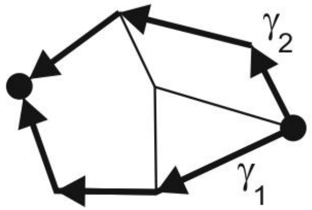
GRW and MERW are equal on regular graphs, but e.g. on defected 2D lattice:

**Generic Random Walk (GRW):** assume uniform distribution among "the nearest neighbors"




**Maximal Entropy Random Walk (MERW):** choose that for each two vertices, each path of given length between them is equally probable



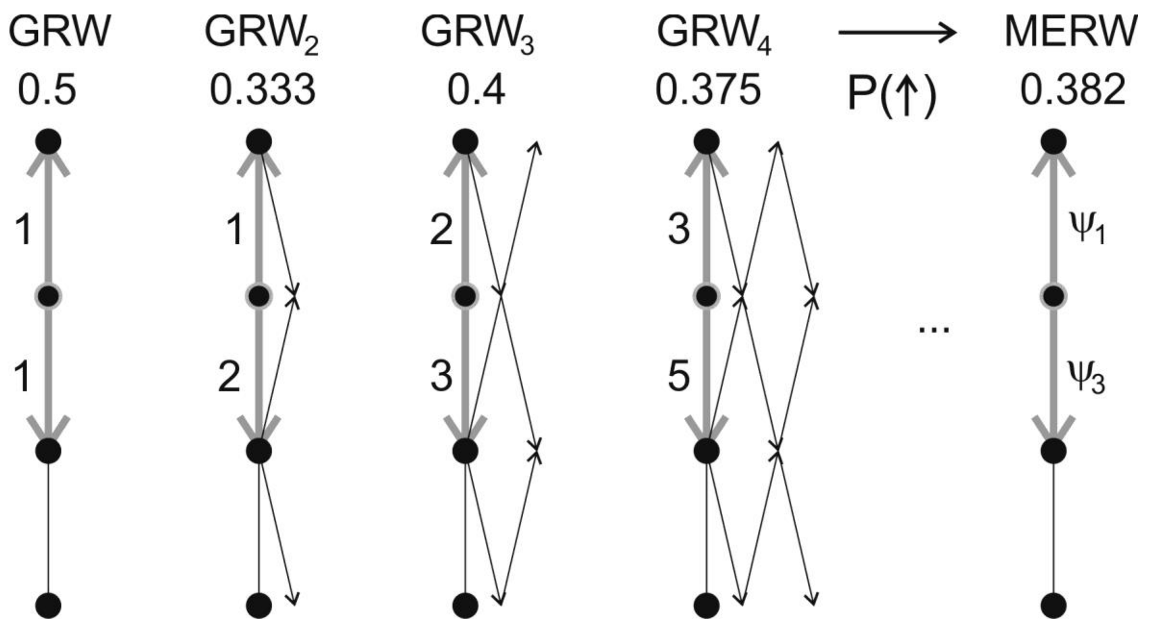

**GRW** assumes we know exactly the used probabilistic algorithm, has characteristic length

**MERW** assumes only there are no hidden local probabilistic rules, is scale-free limit of GRW

## MERW as scale-free limit of GRW

$$S_{ab}^{GRW_k} \propto M_{ab} \sum_c (M^{k-1})_{bc}$$

GRW: each outgoing **edge** is equally probable $(k = 1)$




$\mathrm{GRW}_k$ – each outgoing **length** $k$ **path is equally probable.**

In the limit, the number of paths starting with $a \to b$ **is proportional** to coordinate $(\psi_b)$ of the **dominant eigenvector** of $M$ :

$$M\psi = \lambda\psi$$

**Frobenius-Perron theorem** for connected graph: real, nondegenerated $\lambda > 0$ , $\forall_a \, \psi_a > 0$

**Normalization** for vertex $a$: $\sum_b M_{ab}\psi_b = (M\psi)_a = \lambda\psi_a$

Finally: **while being in $a$, probability of jumping to $b$ is:**
(symmetric $M$:)

$$S_{ab} = \frac{M_{ab}}{\lambda}\frac{\psi_b}{\psi_a}$$

nonlocality

For which stationary probability distribution $(\pi S = \pi)$ is $\pi_a \propto \psi_a^2$

$$(\pi S)_b = \sum_a \psi_a^2 \cdot \frac{M_{ab}}{\lambda}\frac{\psi_b}{\psi_a} = \sum_a \psi_a M_{ab} \cdot \frac{\psi_b}{\lambda} = \lambda\psi_b \frac{\psi_b}{\lambda} = \psi_b^2 = \pi_b$$

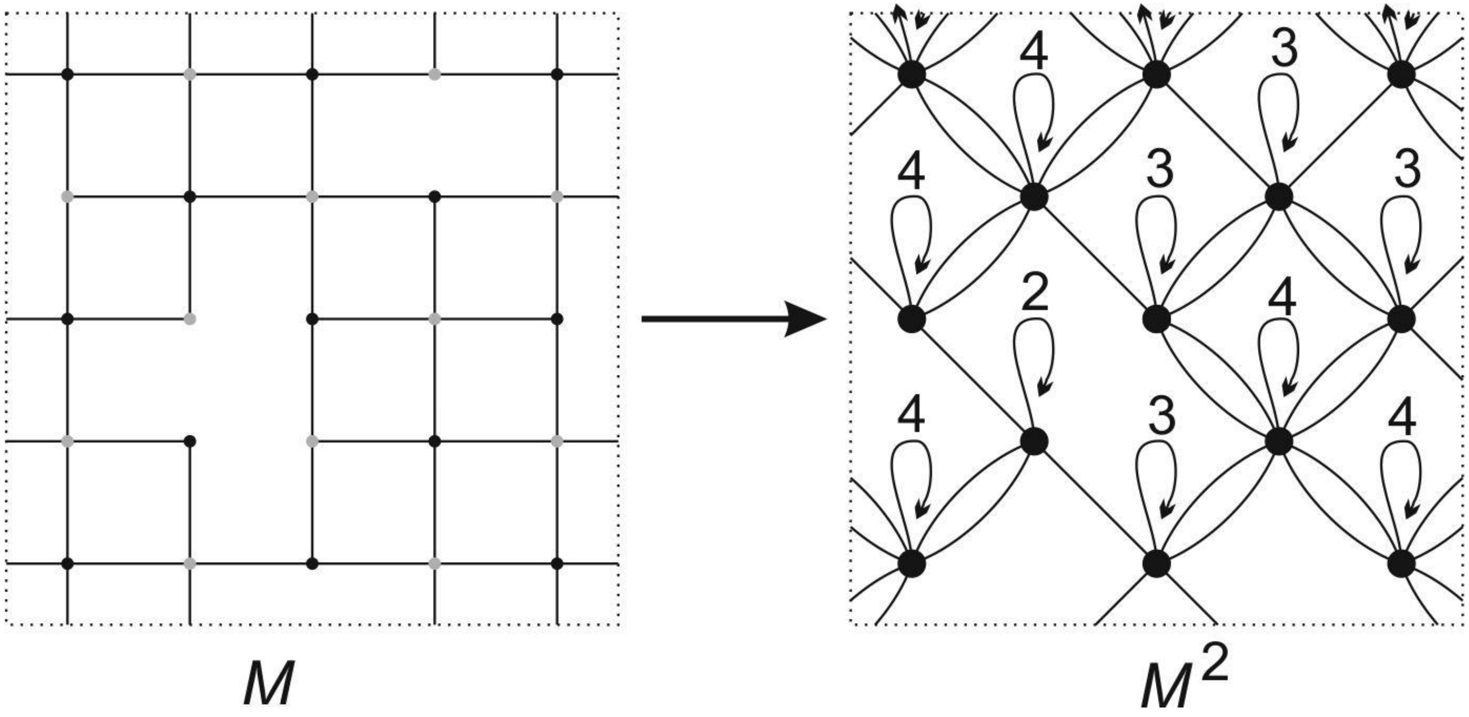
$$(S^k)_{ab} = \frac{(M^k)_{ab}}{\lambda^k}\frac{\psi_b}{\psi_a}$$

# **Renormalization (**being **scale-free)**

We can change not only time scale, but also spatial

$$\left(\left(S^{\mathrm{MERW}(M)}\right)^l\right)_{ij} = \sum_{\gamma_1,\ldots,\gamma_{k-1}} \frac{M_{i\gamma_1}}{\lambda}\frac{\psi_{\gamma_1}}{\psi_i}\cdot\frac{M_{\gamma_1\gamma_2}}{\lambda}\frac{\psi_{\gamma_2}}{\psi_{\gamma_1}}\cdot\ldots\cdot\frac{M_{\gamma_{k-1}\gamma_k}}{\lambda}\frac{\psi_j}{\psi_{\gamma_{k-1}}} = \frac{(M^l)_{ij}}{\lambda^k}\frac{\psi_{\gamma_k}}{\psi_{\gamma_0}} = \left(S^{\mathrm{MERW}(M^l)}\right)_{ij}$$

Usually not true for GRW



# Approximating MERW for short range knowledge ($GRW_k$)

Sinatra, R., Gomez-Gardenes, J., Lambiotte, R., Nicosia, V. & Latora, V. *Maximal-entropy random walks in complex networks with limited information*, Phys. Rev. E 83, 030103 (2011)

| | $GRW$ | $GRW_2$ | $GRW_3$ | $MERW$ |
|---|---|---|---|---|
| | $\frac{h(\pi^0)}{h(\pi)}$ | $\frac{h(\pi^1)}{h(\pi)}$ | $\frac{h(\pi^2)}{h(\pi)}$ | $h_{\max} = h(\pi)$ |
| Regular lattice | 1.000 | 1.000 | 1.000 | 1.79 |
| Random regular graph | 1.000 | 1.000 | 1.000 | 1.79 |
| ER random graph | 0.954 | 0.993 | 0.998 | 1.98 |
| Uncorrelated scale-free $\gamma = 1.5$ | 0.886 | 0.992 | 0.996 | 2.36 |
| BA model | 0.825 | 0.976 | 0.996 | 2.52 |
| Assortative scale-free $\gamma = 1.5$ | 0.876 | 0.991 | 0.999 | 2.44 |
| Disassortative scale-free $\gamma = 1.5$ | 0.937 | 0.990 | 0.997 | 2.18 |
| Regular lattice (1% defects) | 0.996 | 0.997 | 0.998 | 1.38 |
| Regular lattice (10% defects) | 0.967 | 0.978 | 0.981 | 1.34 |
| Regular lattice (20% defects) | 0.931 | 0.955 | 0.963 | 1.29 |
| Internet autonomous system [22] | 0.744 | 0.900 | 0.980 | 4.10 |
| U.S. Airports [18] | 0.879 | 0.990 | 0.997 | 3.88 |
| E-mail [23] | 0.881 | 0.983 | 0.997 | 3.03 |
| SCN (cond-mat) [24] | 0.694 | 0.867 | 0.946 | 3.17 |
| SCN (astro-ph) [24] | 0.784 | 0.941 | 0.973 | 4.41 |
| PGP [25] | 0.597 | 0.920 | 0.976 | 3.75 |

$h \to h_{max}$ but the behavior can be qualitatively different

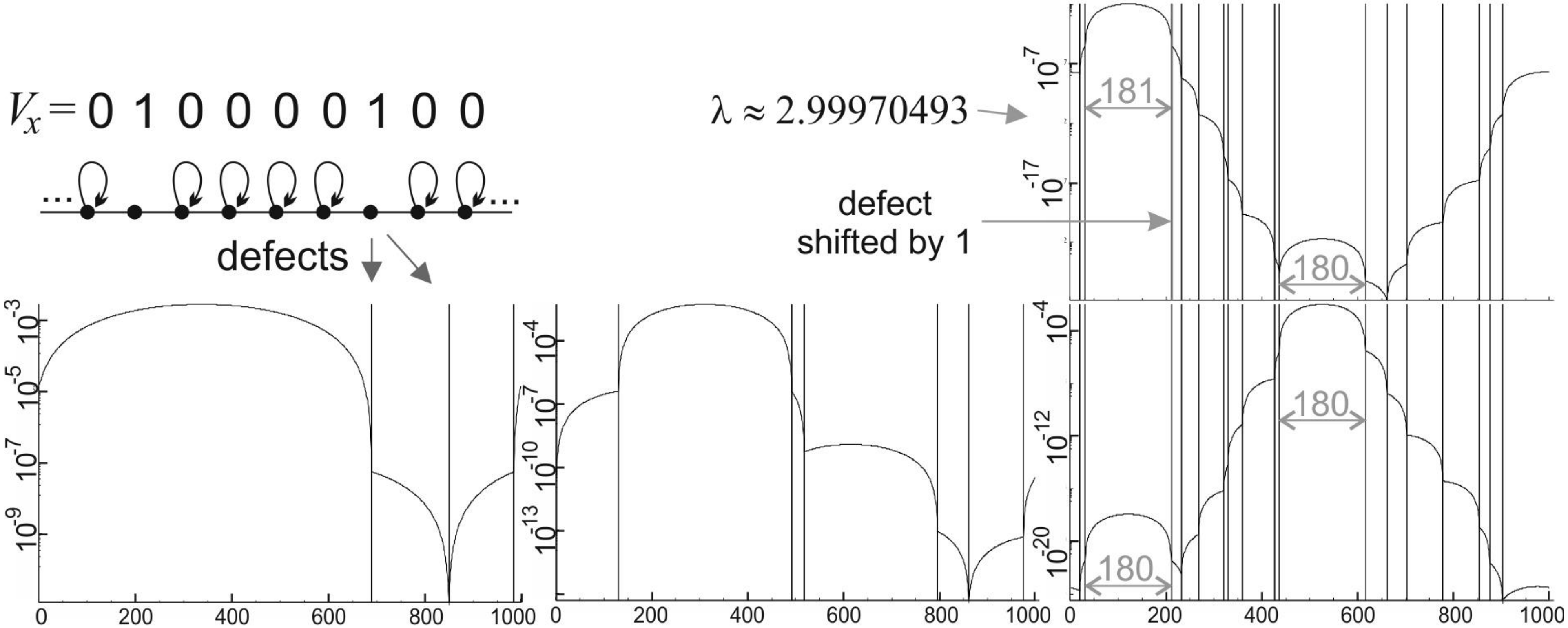
**GRW:** stationary probability $\propto d_i = \sum_j M_{ij}$

**MERW**: stationary probability $\propto \psi^2$ where $M\psi = \lambda\psi$ for largest $\lambda$

Defected $(\lambda\psi)_x = (M\psi)_x = \psi_{x-1} + (1 - V_x)\psi_x + \psi_{x+1}$ $/-3\psi_x$ $/\cdot -1$

1D lattice $E\psi_x = -(\psi_{x-1} - 2\psi_x + \psi_{x+1}) + V_x\psi_x$ for smallest $E = 3 - \lambda$

**Nonlocal** – depends on the whole graph!




$p_d$= 0.002 λ ≈ 2.99998025 $p_d$= 0.01 λ ≈ 2.99992549 $p_d$= 0.03 λ ≈ 2.99970172

## (**diffusion**) A basic question for many complex systems:
# what stationary probability density should we expect?

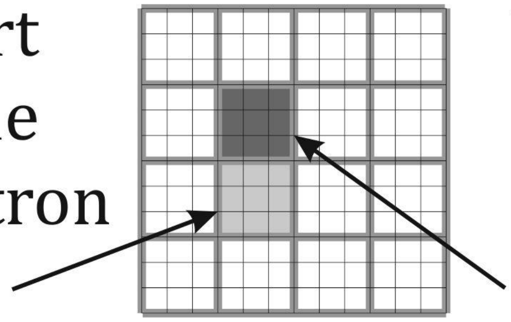
For example for electrons "hopping" between atoms in a lattice

two answers (should agree in applicability intersection):

| Quantum mechanics: | Diffusion: |
|---|---|
| Define energy density for given system: **Hamiltonian** $(\widehat{H})$**,** find its dominant eigenvector/eigenfunction $(\psi)$: | Choose transition probabilities – - stochastic matrix/operator $(\hat{S})$ , and ask for its stationary density: dominant eigenvector/eigenfunction |
| $\widehat{H}\psi = \lambda\psi, \qquad \rho = \|\psi\|^2$ | $\rho\hat{S} = \rho$ |
| Strong localization property (e.g. Anderson's) | Usually weak localization property |

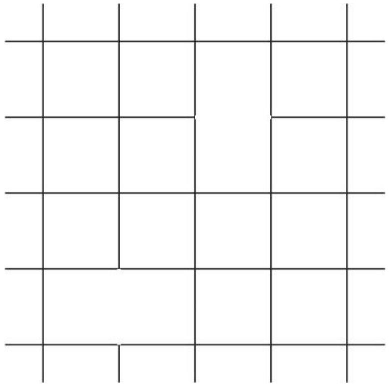
"Stochastic" questions available for **macroscopic** situations: (Heisenberg uncertainty influence microscopic ones)

insert single electron here



What is the probability of finding it here after 1ns ?

Idealized situation: **defected lattice** (cyclic boundary conditions) →



"Natural" stochastic <u>choice</u> ("drunken sailor"):
Each outgoing edge is equally probable (GenericRW)

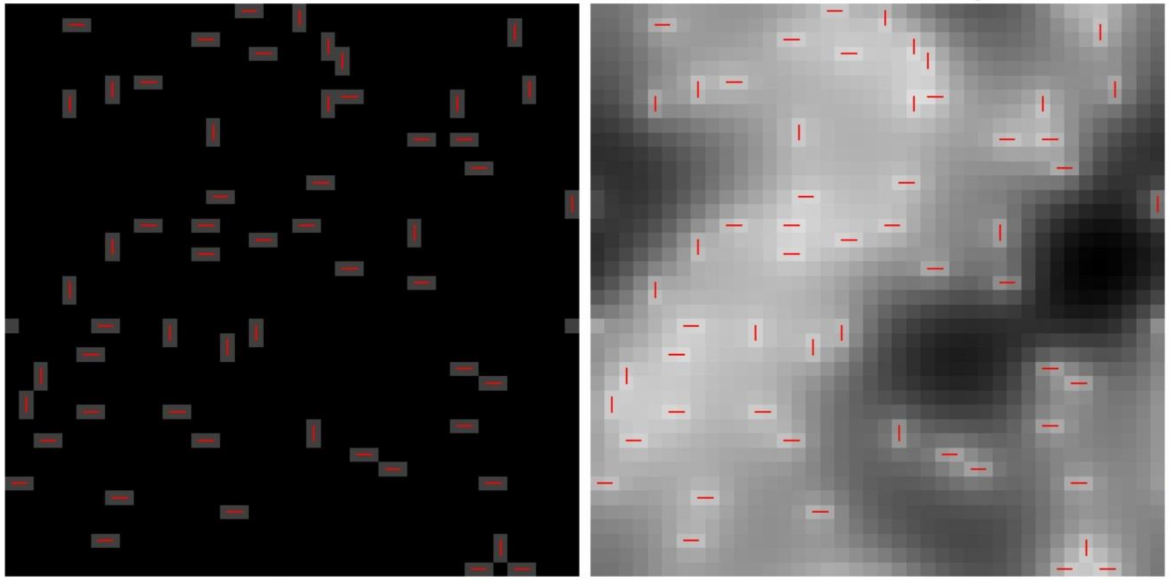
**Bose-Hubbard Hamiltonian** (→ **Schrödinger**) <u>for single particle</u>:

$\hat{H} = -t \sum_{(i,j)\in\mathcal{E}} \left( \hat{a}_j^+ \hat{a}_i + \hat{a}_i^+ \hat{a}_j \right) = -t \cdot$ "adjacency matrix"

STM measurenments of electron density for $Ga_{1-x}Mn_x As$ (20pA)

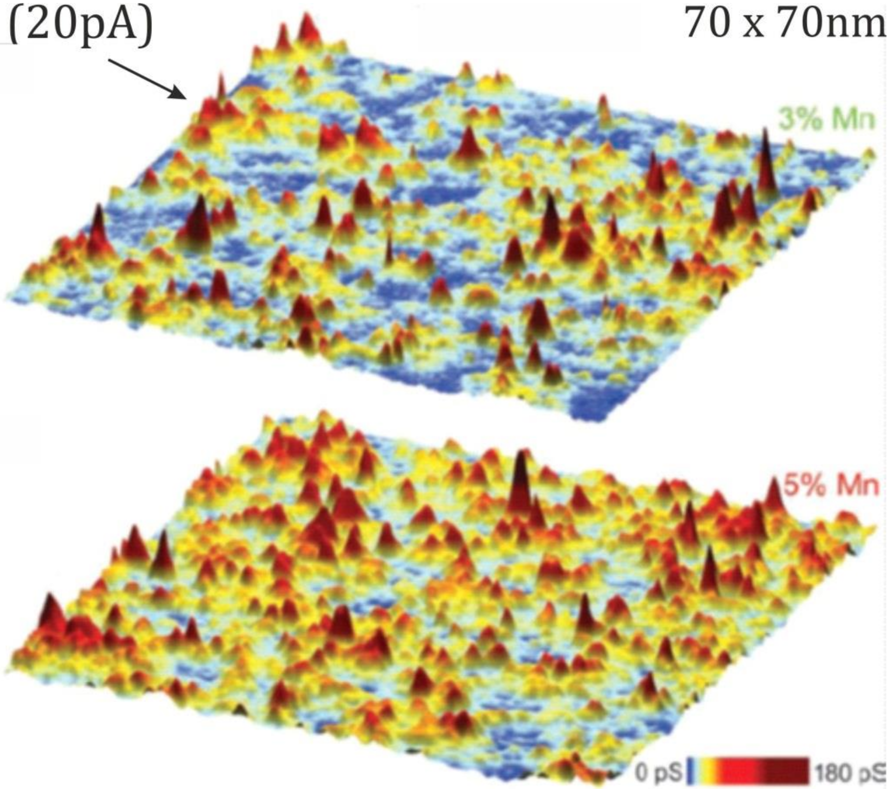
GRW

ground state density of Bose-Hubbard / MERW

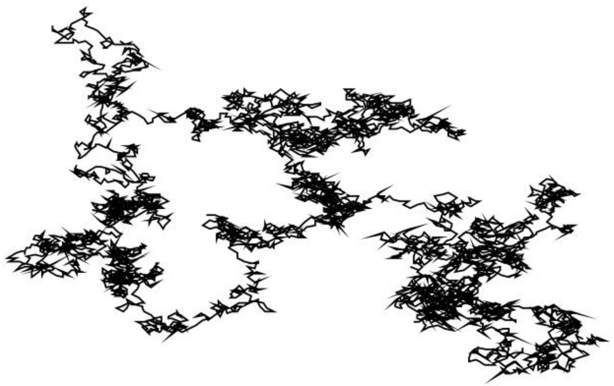


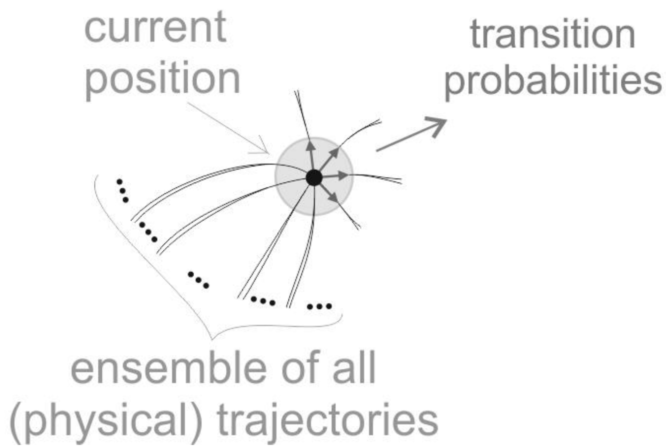
conductor

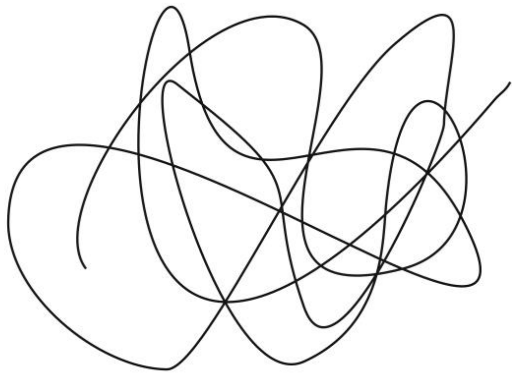
insulator




Discrepancy source: **GRW only approximates maximal uncertainty principle** stochastic picture | thermodynamical picture | ergodic picture





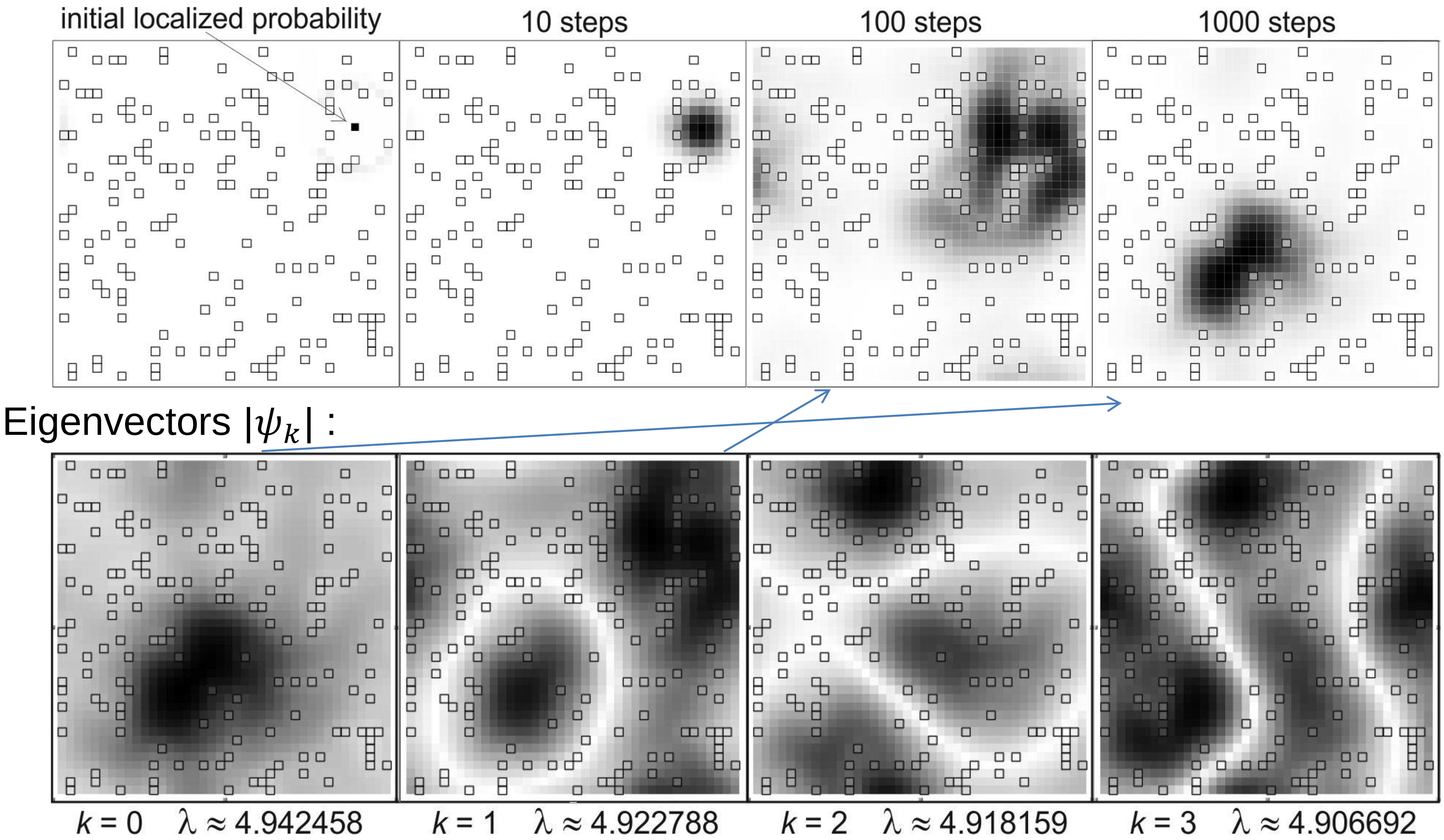
**Stochastic** picture **–** the evolution is indeed succeeding random decisions, accordingly to chosen by us transition probabilities – locally maximizing entropy, no **localization property**
**Ergodic** picture – evolution is usually fully determined, but because of chaotic behavior we introduce densities by averaging over single trajectory (thermodynamical fluctuations?)

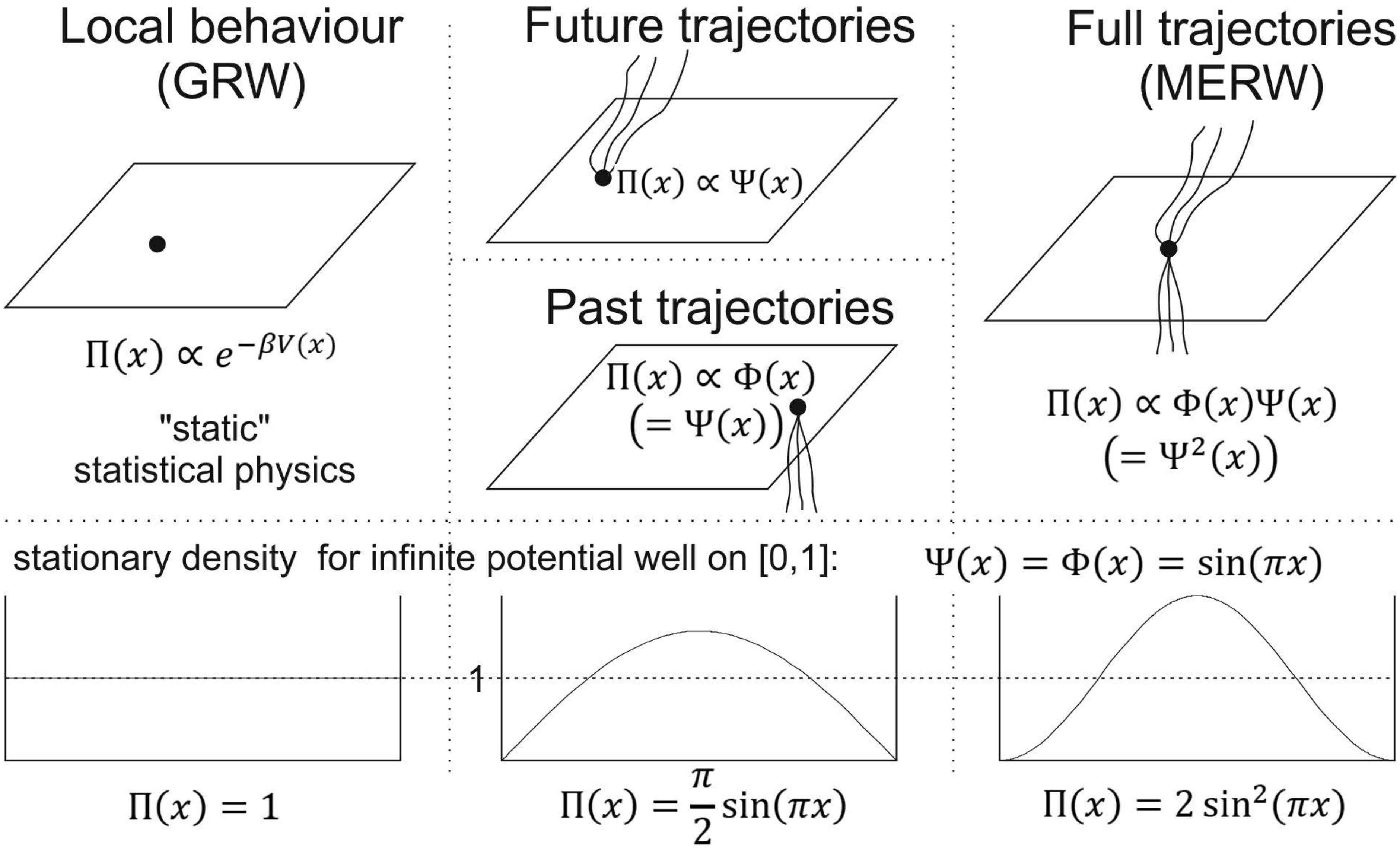
**Thermodynamical** picture**: system too complicated** - use maximal uncertainty principle/canonical ensemble to predict the most probable behavior only.
– transition probabilities calculated from canonical ensemble among possible trajectories going through given point – fully optimizing entropy (free energy),
– object doesn't directly use these probabilities (**nonlocal** - depend on the whole space), but just somehow chooses a trajectory (**not imposing any local probabilistic rules!**)
Only we use the found probabilities to estimate the probability density of its position,
– stationary density has strong localization property – to thermal equilibrium predicted by quantum mechanics – ground state density of corresponding Hamiltonian.

# MERW evolution:

$$\left(S^{\mathrm{M}}\right)^t_{ij} = \frac{(M)^t_{ij}}{\lambda_0^t}\frac{\psi_{0,j}}{\psi_{0,i}} = \left(\sum_k \left(\frac{\lambda_k}{\lambda_0}\right)^t \varphi_{k,j}\psi_{k,i}\right)\frac{\psi_{0,j}}{\psi_{0,i}}$$

First “stochastic shift” toward **near** (overlapping) eigenvectors (sub-diffusion), then “deexcitate” toward nearer **ground state** (super-diffusion)





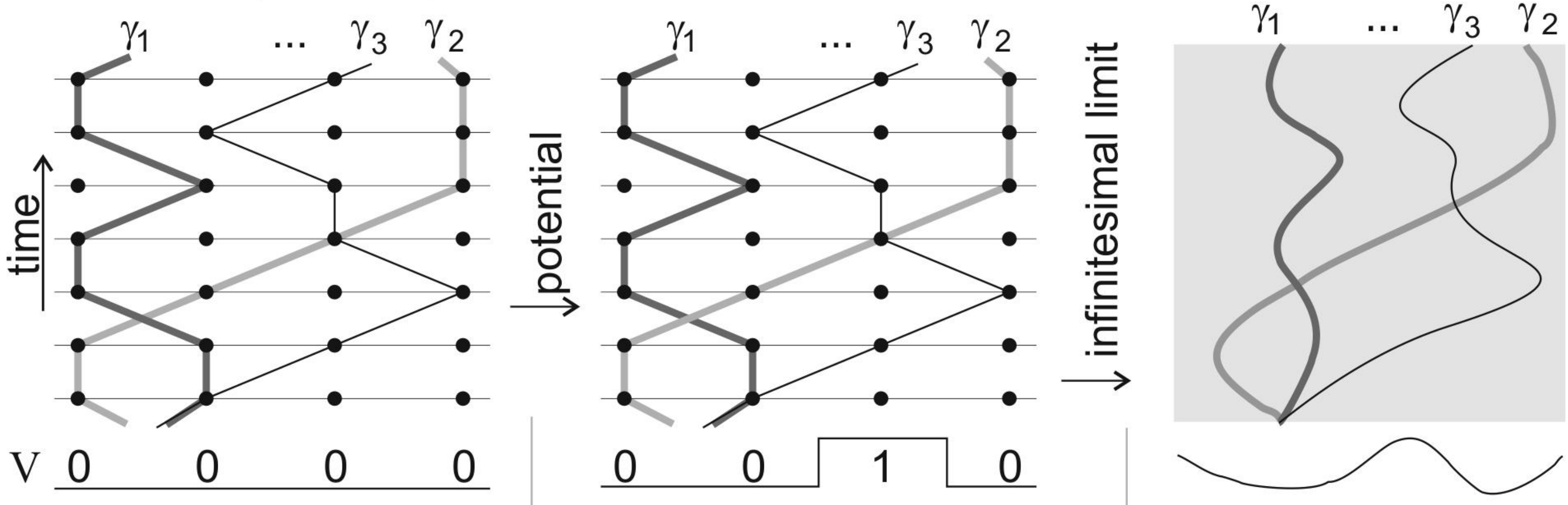

**GRW:** assume some concrete transition probabilities

**MERW**: assume there is no base to assume anything concrete about transition probabilities

Add potential to emphasize some scenarios: **Boltzmann distribution**

maximizes entropy while fixed sum of energies (minimizes free energy)

$$\max_{(p_i):\sum_i p_i=1} \left(\sum_i p_i \ln(1/p_i) - \sum_i p_i E_i\right) = \ln\left(\sum_i e^{-E_i}\right) \quad \text{for} \quad p_i \propto e^{-E_i}$$




eigenvector

| equiprobable paths (MERW) | Boltzmann paths | Boltzmann trajectories |
| --- | --- | --- |
| $P(\gamma_1) = P(\gamma_2) = P(\gamma_3)$ | $P(\gamma_1) > P(\gamma_2) > P(\gamma_3)$ | $P(\gamma_1) > P(\gamma_2) > P(\gamma_3)$ |
| $-(\psi_{i-1} - 2\psi_i + \psi_{i+1}) = E_0\psi_i$ | $-(\psi_{i-1} - 2\psi_i + \psi_{i+1}) + V_i\psi_i \approx E_0\psi_i$ | $-\Delta\Psi + V\Psi = E_0\Psi$ |
| $\pi_i \propto \psi_i^2$ | $\pi_i \propto \psi_i^2$ | $\Pi(x) \propto \Psi^2(x)$ |

energy of path $(\gamma_t\gamma_{t+1} \ldots \gamma_s)$ is $V_{\gamma_t\gamma_{t+1}} + \cdots + V_{\gamma_{s-1}\gamma_s}$

**Boltzmann distribution** among paths – use matrix: $M_{ij} = A_{ij} e^{-\beta V_{ij}}$

$$S_{\gamma_0\gamma_1} S_{\gamma_1\gamma_2} \ldots S_{\gamma_{l-1}\gamma_l} = \frac{M_{\gamma_0\gamma_1} \ldots M_{\gamma_{l-1}\gamma_l}}{\lambda^l} \frac{\psi_{\gamma_l}}{\psi_{\gamma_0}} = \frac{e^{-\beta\left(V_{\gamma_0\gamma_1}+V_{\gamma_1\gamma_2}+..+V_{\gamma_{l-1}\gamma_l}\right)}}{\lambda^l} \frac{\psi_{\gamma_l}}{\psi_{\gamma_0}}$$

---

**Eigenequation for 1D lattice:** $\epsilon$ – time step, $\delta$ – lattice constant

$$\lambda_\epsilon \psi_i = (M_\epsilon \psi)_i = e^{-\beta\epsilon\frac{V_{i-1}+V_i}{2}} \psi_{i-1} + e^{-\beta\epsilon V_i} \psi_i + e^{-\beta\epsilon\frac{V_i+V_{i+1}}{2}} \psi_{i+1}$$

$$\lambda_\epsilon \psi_i \approx \psi_{i-1} + \psi_i + \psi_{i+1} - \epsilon\beta\left(\frac{V_{i-1}+V_i}{2}\psi_{i-1} + V_i\psi_i + \frac{V_i+V_{i+1}}{2}\psi_{i+1}\right)$$

$$\lambda_\epsilon \psi_i \approx \psi_{i-1} + \psi_i + \psi_{i+1} - 3\epsilon\beta V_i \psi_i \qquad /-3\psi_i \qquad /\cdot \frac{-1}{3\beta\epsilon}$$

$$\frac{3-\lambda_\epsilon}{3\beta\epsilon}\psi_i \approx -\frac{1}{3\beta}\frac{\psi_{i-1} - 2\psi_i + \psi_{i+1}}{\epsilon} + V_i\psi_i$$

$\epsilon \to 0$

$$\epsilon = \frac{\delta^2}{3\alpha}, \qquad E_\epsilon = \frac{3-\lambda_\epsilon}{3\beta\epsilon} \longrightarrow \boxed{E\Psi = \left(-\frac{\alpha}{\beta}\Delta + V\right)\Psi}$$

Going to normalized $\Psi^2(x)$ stationary probability density for the lowest possible $E$

Propagator: $$S^t(x,y) = \frac{\langle x|e^{-t\beta\hat{H}}|y\rangle}{e^{-t\beta E_0}} \frac{\Psi_0(y)}{\Psi_0(x)} = \frac{\sum_i e^{-t\beta E_i}\langle x|\Psi_i\rangle\langle\Psi_i|y\rangle}{e^{-t\beta E_0}} \frac{\Psi_0(y)}{\Psi_0(x)}$$

**Time dependence** – e.g. potential can vary with time: $M_{ij}^t = A_{ij} e^{-\beta V_{ij}^t}$

energy of path $(\gamma_t \gamma_{t+1} \dots \gamma_s)$ is $V_{\gamma_t \gamma_{t+1}}^t + \dots + V_{\gamma_{s-1}\gamma_s}^{s-1}$ where $V_{ij}^t \equiv V_{ij}(t)$

Generalized dominant eigenvectors: density on the end of **past** and **future** ensembles

$$\varphi_j^t := \lim_{l\to\infty} \frac{\sum_i (M^{t-l} M^{t-l+1} \dots M^{t-1})_{ij}}{\tilde{N}^t(l)} \qquad \psi_i^t := \lim_{l\to\infty} \frac{\sum_j (M^t M^{t+1} \dots M^{t+l-1})_{ij}}{N^t(l)} \qquad (\geq 0)$$

$$((\varphi^t)^T M^t)_j = \lim_{l\to\infty} \frac{\sum_i (M^{t-l} M^{t-l+1} \dots M^t)_{ij}}{\tilde{N}^t(l)} = \tilde{\lambda}^t \varphi_j^{t+1} \quad \text{where} \quad \tilde{\lambda}^t = \lim_{l\to\infty} \frac{\tilde{N}^{t+1}(l+1)}{\tilde{N}^t(l)}$$

$$(M^t \psi^{t+1})_i = \lim_{l\to\infty} \frac{\sum_i (M^t M^{t+1} \dots M^{t+l})_{ij}}{\tilde{N}^{t+1}(l)} = \lambda^t \psi_i^t \quad \text{where} \quad \lambda^t = \lim_{l\to\infty} \frac{N^t(l+1)}{N^{t+1}(l)}$$

Stationary probability: $p_i^t = \varphi_i^t \psi_i^t$ propagator $S_{ij} = \frac{M_{ij}^t}{\lambda^t} \frac{\psi_j^{t+1}}{\psi_j^t}$

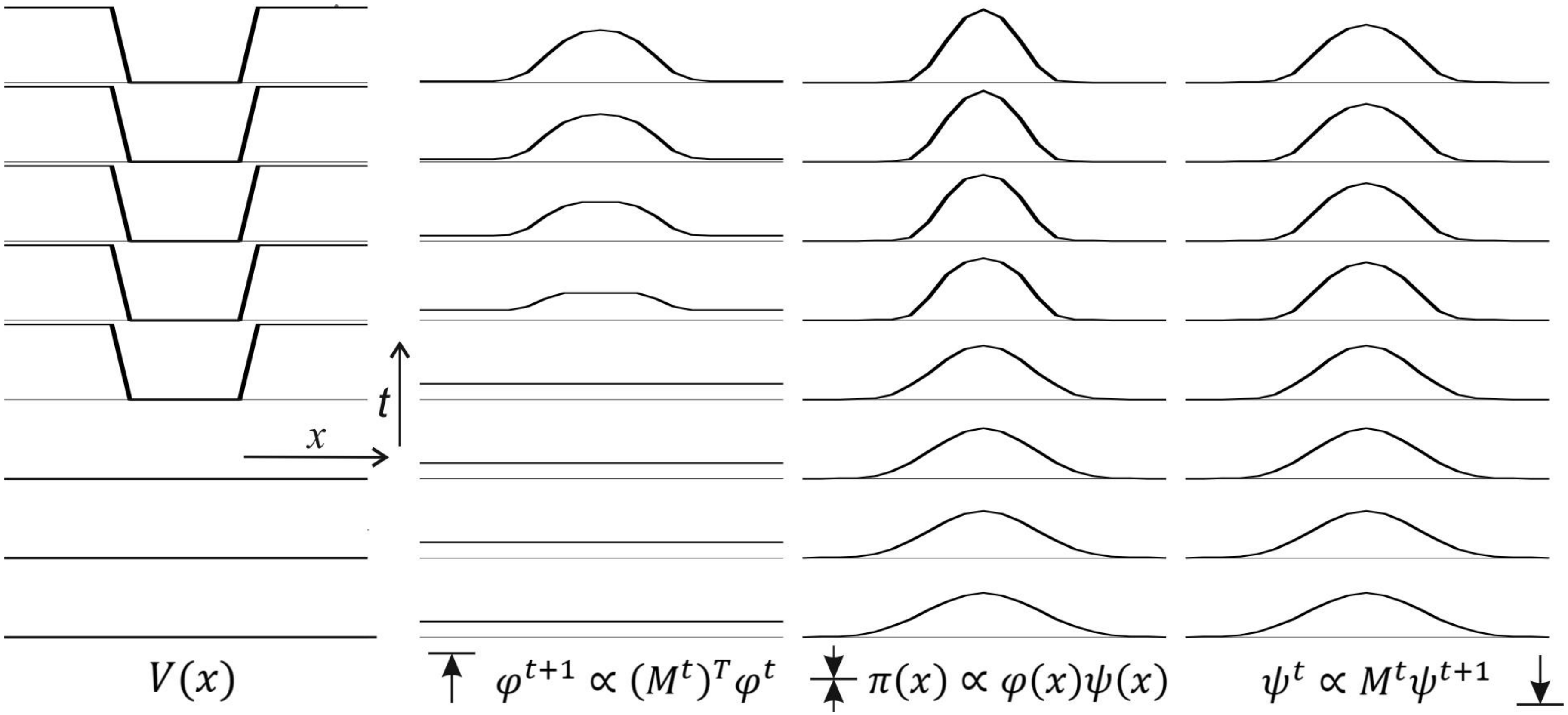
$$(S^{ts})_{ij} := (S^t S^{t+1} \dots S^{s-1})_{ij} = \frac{(M^t M^{t+1} \dots M^{s-1})_{ij}}{\lambda^t \lambda^{t+1} \dots \lambda^{s-1}} \frac{\psi_j^s}{\psi_i^t}$$

Conserved probability? $(\varphi^t)^T \psi^t = (\varphi^t)^T \frac{M^t \psi^{t+1}}{\lambda^t} = (\varphi^t)^T M^t \frac{\psi^{t+1}}{\lambda^t} = \frac{\tilde{\lambda}^t}{\lambda^t} (\varphi^{t+1})^T \psi^{t+1}$

**Continuity equation** $\Leftrightarrow \lambda = \tilde{\lambda}$ (exact values only balance between $\varphi$ and $\psi$)

Final evolution equation: $\lambda^t \varphi^{t+1} = (M^t)^T \varphi^t$ $\quad M^t \psi^{t+1} = \lambda^t \psi^t$

**Adiabatic approximation**: If $V$ is locally constant, $\varphi$ and $\psi$ are approximately right and left dominant eigenvectors of $M$ … but generally:




**Boltzmann distribution among paths is time-symmetric**

It is effective model: only represents our knowledge
We know about the change – that later particle should be in the well, so earlier it should be nearby

Now for **1D lattice** there appears additional time derivative:

$$\lambda_\epsilon^t \psi_x^t = (M^t \psi^{t+1})_x \approx \psi_{x-1}^{t+1} + \psi_x^{t+1} + \psi_{x+1}^{t+1} - 3\epsilon\beta V_x^t \psi_x^t \qquad /-3\psi_x^{t+1} \qquad /\cdot \frac{-1}{3\epsilon\beta}$$

$$\frac{1}{\beta}\,\frac{\psi_x^{t+1}-\psi_x^t}{\epsilon} - E_\epsilon^t \psi_x^t \approx -\frac{1}{3\beta}\frac{\psi_{x-1}^{t+1}-2\psi_x^{t+1}+\psi_{x+1}^{t+1}}{\epsilon} + V_x^t\psi_x^t \qquad \text{for} \qquad E_\epsilon^t := \frac{3-\lambda_\epsilon^t}{3\epsilon\beta}$$

Finally choosing $\epsilon = \frac{\delta^2}{3\alpha}$ in infintesimal limit we get **evolution equations**:

$$\frac{d}{dt}\Phi = \beta\left(E - \hat{H}\right)\Phi \qquad \frac{d}{dt}\Psi = \beta\left(\hat{H} - E\right)\Psi \qquad \text{for} \qquad \hat{H} = -\frac{\alpha}{\beta}\Delta + V$$

$\Phi$ should evolve forward in time (to be stable), $\Psi$ backward

In **adiabatic approximation** $\Phi \approx \Psi$ for $E(t) = \langle \Phi(t) | \hat{H}(t) \Psi(t) \rangle$

---

$$\frac{d}{dt}(\Phi\Psi) = \beta\left(\left(\left(E-\hat{H}\right)\Phi\right)\Psi + \Phi\left(\hat{H}-E\right)\Psi\right) = \alpha\left((\Delta\Phi)\Psi - \Phi(\Delta\Psi)\right) = \alpha\nabla\cdot((\nabla\Phi)\Psi - \Phi(\nabla\Psi))$$

**Continuity equation**: $\frac{d}{dt}\rho = -\nabla\cdot J$ for $J = \alpha(\Phi\nabla\Psi - \Psi\nabla\Phi)$

Quantum ($\psi \in \mathbb{C}$) : $j = \frac{\hbar}{2mi}(\bar{\psi}\nabla\psi - \psi\nabla\bar{\psi})$ substituting $\psi = \frac{e^{i\gamma}}{\sqrt{2}}(\Phi + i\Psi)$

$$j = \frac{\hbar e^{i\gamma} e^{-i\gamma}}{4mi}\left((\Phi - i\Psi)\nabla(\Phi + i\Psi) - (\Phi + i\Psi)\nabla(\Phi - i\Psi)\right) = \frac{\hbar}{2m}(\Phi\nabla\Psi - \Psi\nabla\Phi)$$

Suggesting to choose $\alpha = \frac{\hbar}{2m}$ $\beta = \frac{2m}{\hbar^2}\alpha = \frac{1}{\hbar}$

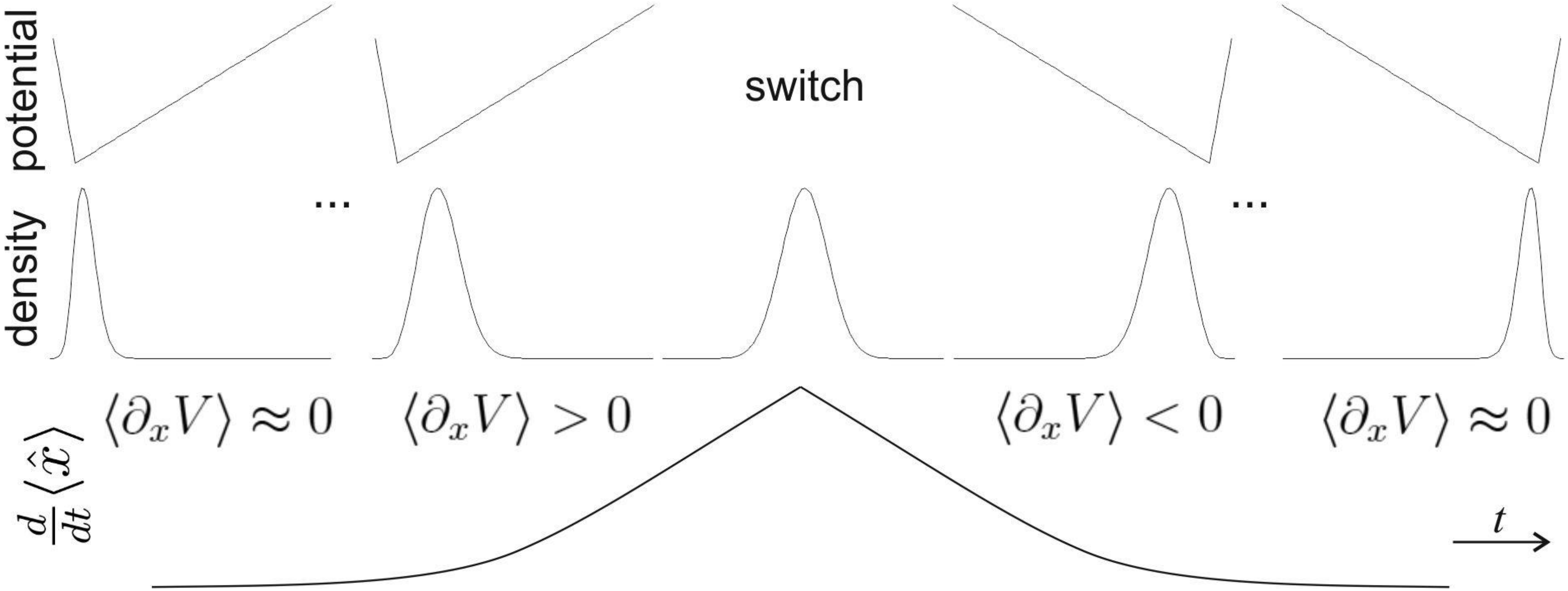
$$\frac{d}{dt}\langle\Phi\big|\hat{O}\Psi\rangle = \beta\langle\Phi\big|\big(E-\widehat{H}\big)\hat{O}\Psi\rangle + \langle\Phi|\frac{\partial\hat{O}}{\partial t}\Psi\rangle + \beta\langle\Phi|\hat{O}\big(\widehat{H}-E\big)\Psi\rangle$$

**Ehrenfest equation:** $\langle\hat{O}\rangle = \langle\frac{\partial\hat{O}}{\partial t}\rangle + \beta\langle\big[\hat{O},\widehat{H}\big]\rangle$

$\big[\hat{x},\widehat{H}\big] = 2\frac{\alpha}{\beta}\nabla \quad \Rightarrow \quad \frac{\mathrm{d}\langle\hat{x}\rangle}{\mathrm{dt}} = \langle 2\alpha\nabla\rangle = \frac{\langle\hat{p}\rangle}{\mathrm{m}}$ for $\hat{p} = 2\alpha\nabla = \hbar\nabla$

Now $\big[\hat{p},\widehat{H}\big] = [\hbar\nabla, V] = \hbar\nabla V \quad \Rightarrow \quad \frac{d}{dt}\langle\hat{p}\rangle = \beta\langle\hbar\nabla V\rangle = \langle\nabla V\rangle = \int\rho(x)\nabla V(x)dx$

Getting opposite than expected: $m\frac{d^2}{dt^2}\langle\hat{x}\rangle = \langle\nabla V\rangle$



In quantum mechanics $\psi$ is complex function
$\langle\psi|\psi\rangle = \text{const}$ because $\langle\psi| \to e^{i\hat{H}t/\hbar}\langle\psi|$ while $|\psi\rangle \to e^{-i\hat{H}t/\hbar}|\psi\rangle$

In MERW $\Phi$ and $\Psi$ are real nonnegative functions
$\langle\Phi|\Psi\rangle$ =const because $\langle\Phi| \to e^{-\beta t(\hat{H}-E)}\langle\Phi|$ while $|\Psi\rangle \to c^{\beta t(\hat{H}-E)}|\Psi\rangle$

This time momentum operator is not self-adjoined:

$$\hat{p} = \hbar\nabla \qquad\qquad \hat{p}^\dagger = -\hbar\nabla$$

$\hat{p}^2$ also is not self-adjoined, so we have to use $\hat{p}^\dagger\hat{p}$ instead

$$\hat{H} = -\frac{\hbar^2}{2m}\Delta + V = \frac{\hat{p}^\dagger\hat{p}}{2m} + V$$

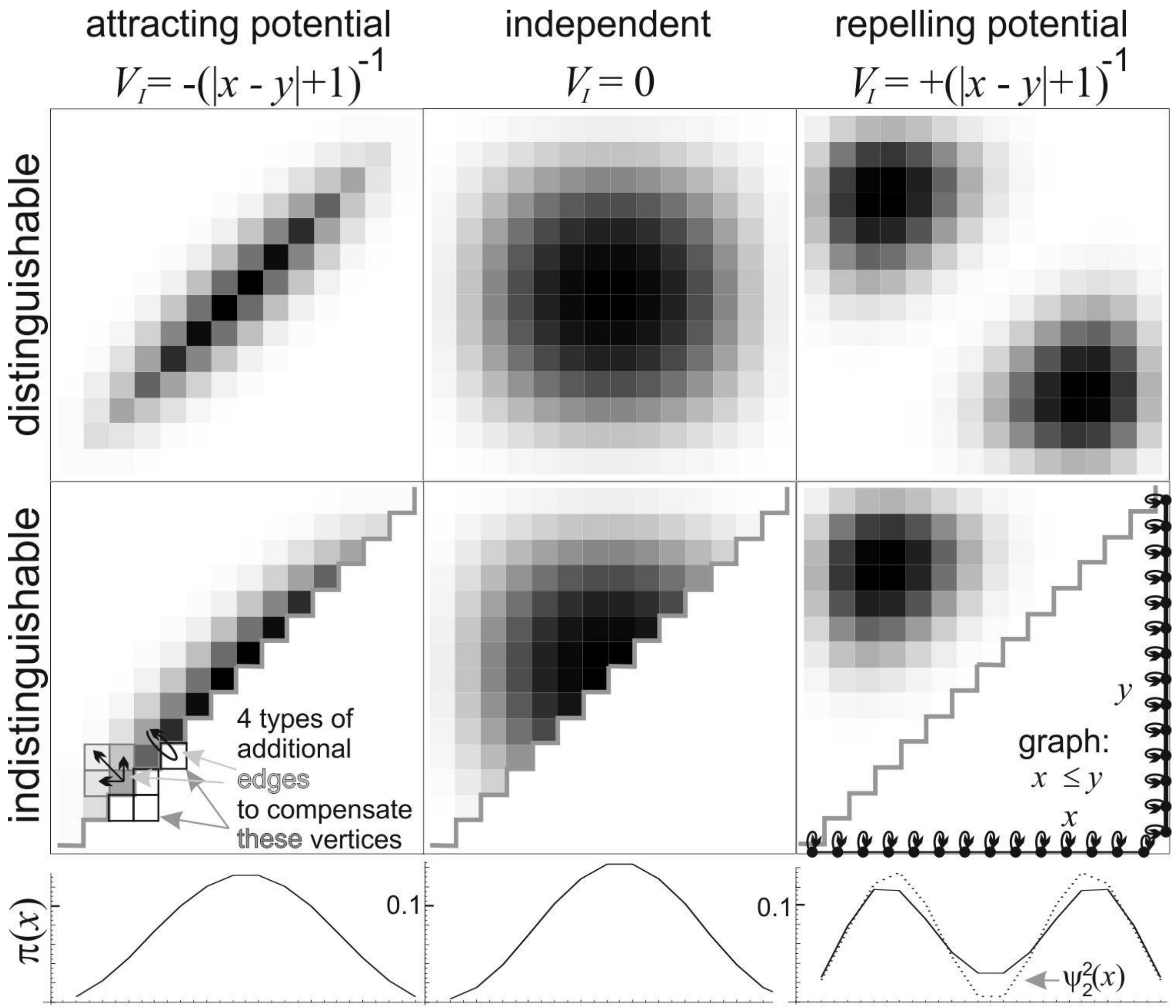
For adiabatic approximation $(\Phi = \Psi)$ we get **Heisenberg principle** analogue:

$$0 \le \langle(\hat{x} + \lambda\hat{p})\Psi|(\hat{x} + \lambda\hat{p})\Psi\rangle = \langle\Psi|(\hat{x} - \lambda\hat{p})(\hat{x} + \lambda\hat{p})\Psi\rangle = \langle\hat{x}^2\rangle + \lambda^2\langle\hat{p}^\dagger\hat{p}\rangle - \lambda\hbar$$

Discriminant $\le 0$:

$$\sqrt{\langle\hat{x}^2\rangle}\sqrt{\langle\hat{p}^\dagger\hat{p}\rangle} \ge \frac{\hbar}{2}$$

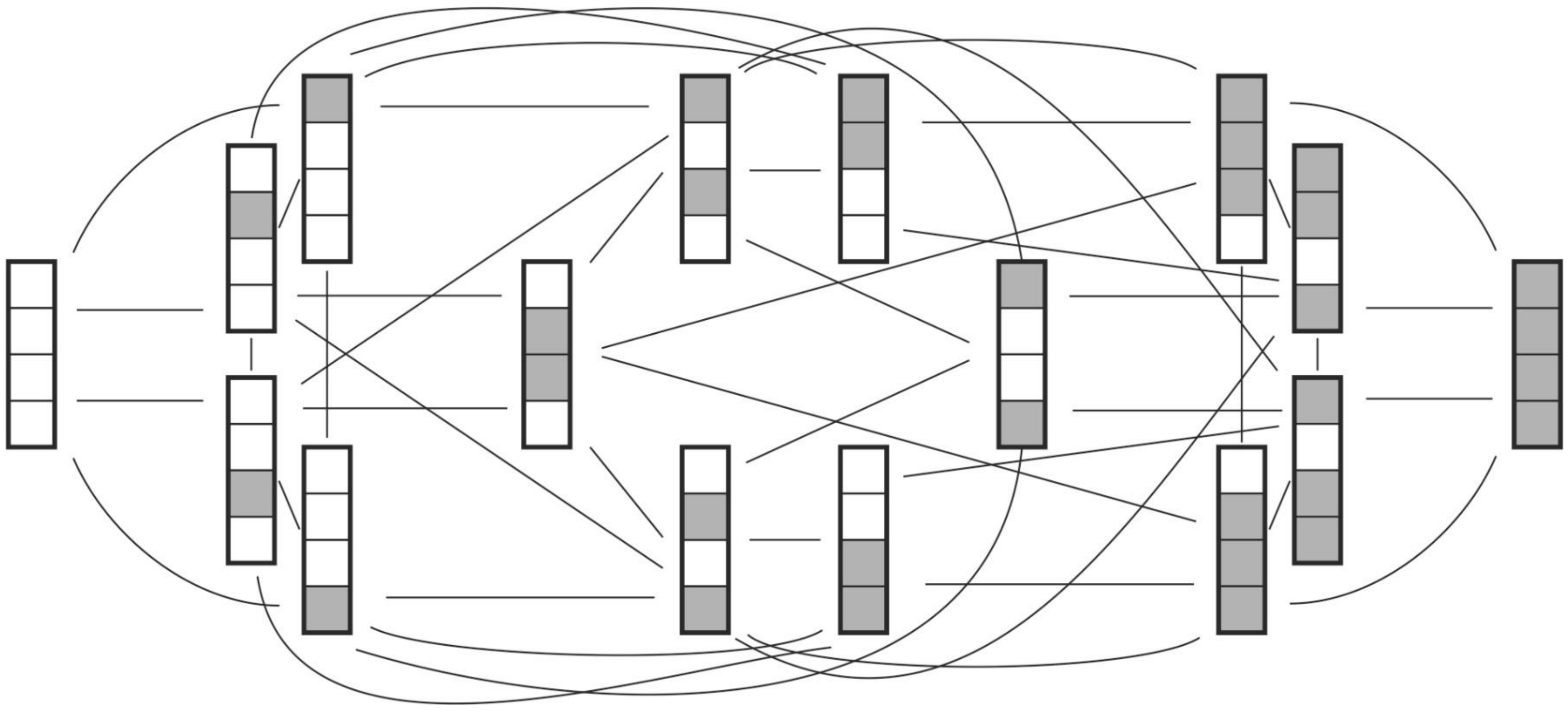
**Two particles** – consider trajectory in the space of pair configurations




Thermodynamical **Pauli exclusion principle**:
repelling particles choose separate dynamical equilibrium **states**

**Various number of particles**: vertex ≡ configuration

For example adjacency matrix for fermions on length 4 segment graph



$|\bar{n}\rangle$ - sum of all $n!$ permutations

$\hat{a}|\bar{n}\rangle = n|\overline{n-1}\rangle$ $\quad$ $\hat{a}^\dagger|\overline{n-1}\rangle = |\bar{n}\rangle$ $\quad$ $\hat{a}^\dagger\hat{a}|\bar{n}\rangle = n|\bar{n}\rangle$ $\quad$ $[\hat{a}, \hat{a}^\dagger] = 1$

Standard normalization: $|n\rangle = |\bar{n}\rangle/\sqrt{n!}$

$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle$ $\quad$ $\hat{a}^\dagger|n-1\rangle = \sqrt{n}|n\rangle$ $\quad$ $\left(\hat{a}^\dagger\right)^n|0\rangle = \frac{1}{\sqrt{n!}}|n\rangle$

**Bose-Hubbard model** – repulsing bosons on lattice

$$\hat{H}_{BH} = -t \sum_{(i,j)\in\mathcal{E}} \hat{a}_j^\dagger \hat{a}_i + \frac{U}{2}\sum_{i\in\mathcal{V}} \hat{n}_i(\hat{n}_i - 1) \qquad \ldots + \sum_{i\in\mathcal{V}} V(i)\hat{n}_i + \sum_{i,j\in\mathcal{V}} V_I(i,j)\hat{n}_i\hat{n}_j$$

Accordingly to **MERW**: diagonal terms ≡ self-loops ("paying for staying")

$$\hat{H}_{MERW} \propto -\sum_{(i,j)\in\mathcal{E}} \hat{a}_j^\dagger \hat{a}_i \, e^{-\epsilon\beta V(\text{configuration before and after transition})} \approx$$

$$\approx -\sum_{(i,j)\in\mathcal{E}} \hat{a}_j^\dagger \hat{a}_i + \epsilon\beta d \sum_{i\in\mathcal{V}} V(\text{configuration after transition})\, \hat{a}_i^\dagger \hat{a}_i$$

Three $\epsilon$ order approximations used exactly as for lattices: $e^{-\epsilon\beta V} \approx 1 - \epsilon\beta V$, that for neighboring vertices, $V$ and coordinates of dominant eigenvector are nearly equal $\left(\hat{a}_i^\dagger \hat{a}_j \approx \hat{a}_i^\dagger \hat{a}_i\right)$.

Both Hamiltonians are practically equivalent for **single particle without potential** and **in continuous limit**, but generally they only approximate each other.

Another question: why only one particle can transit at once?

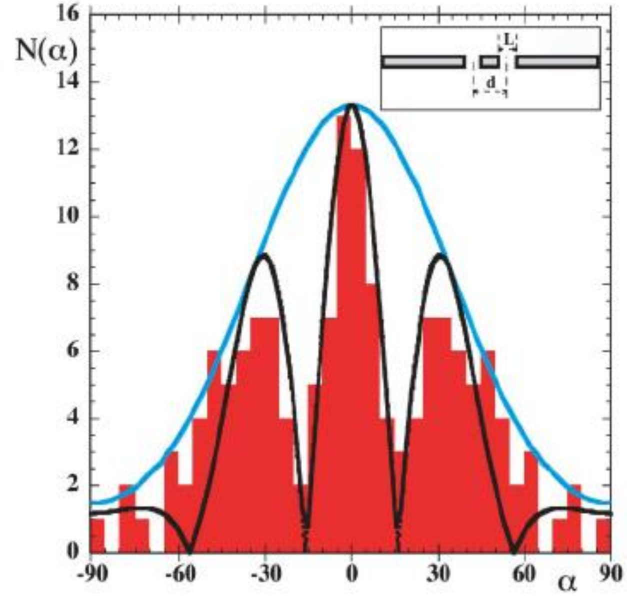
# **Macroscopic soliton** model – **oil droplet** maintaining shape due to surface tension

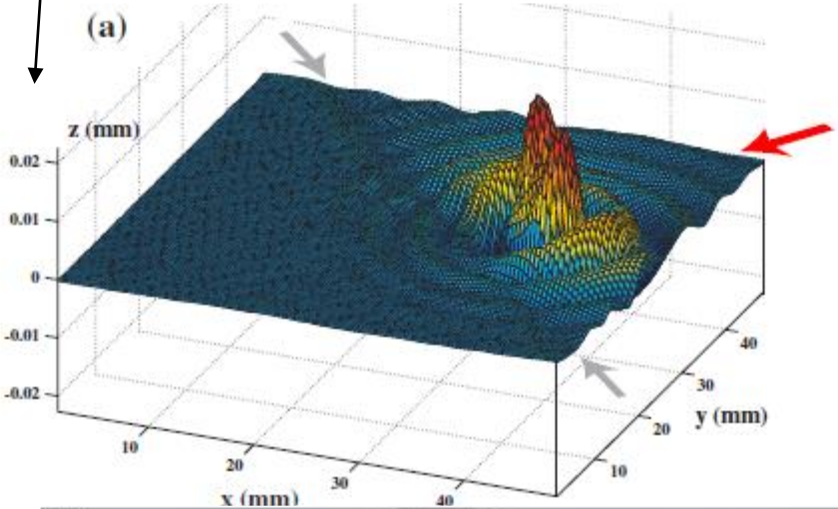
**Bouncing droplet** on vertically vibrated bath is **coupled to the surface waves it generates**. Becomes a "walker" moving at **constant velocity**.

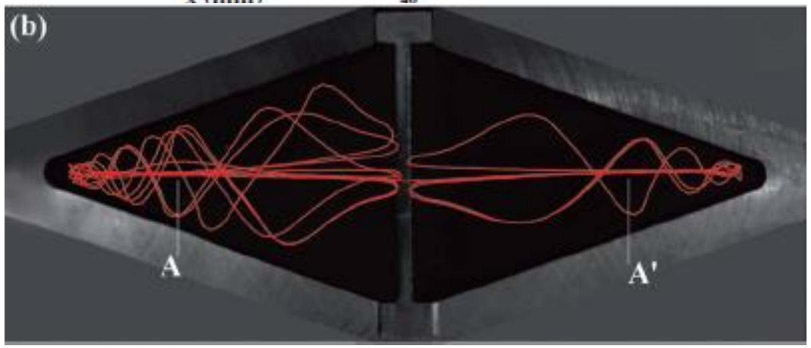
Y. Couder and E. Fort, **Single-Particle Diffraction and Interference at a Macroscopic Scale,** Phys. Rev. Lett. 97 (2006)

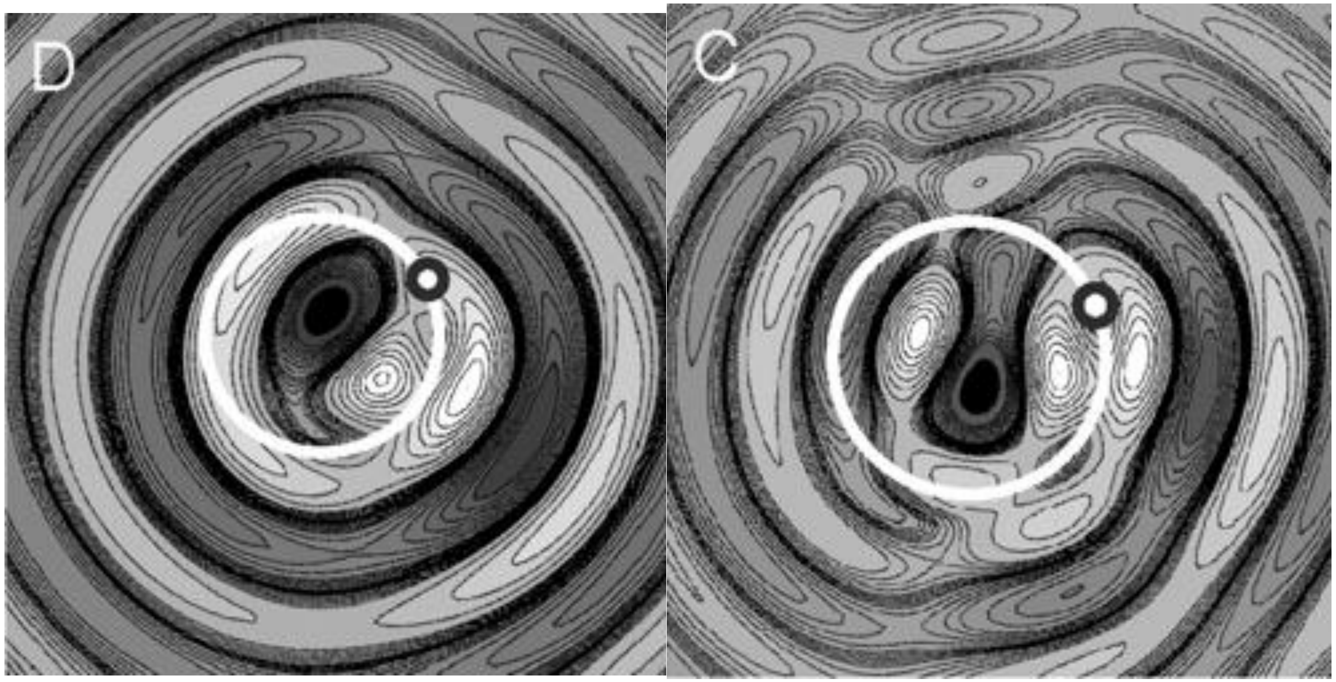
A. Eddi, E. Fort, F. Moisy, and Y. Couder, **Unpredictable Tunneling of a Classical Wave-Particle Association**, Phys. Rev. Lett. 102 (2009)

E. Fort, A. Eddib, A. Boudaoudc, J. Moukhtarb, and Y. Couderb, **Path-memory induced quantization of classical orbits,** PNAS vol. 107 (2010)



**Summary of diffusion part:** If instead of guessing the stochastic propagator (assuming that the walker indeed uses these probabilities), we assume the maximal uncertainty principle (only we use these probabilities), the predictions are no longer in disagreement with QM.

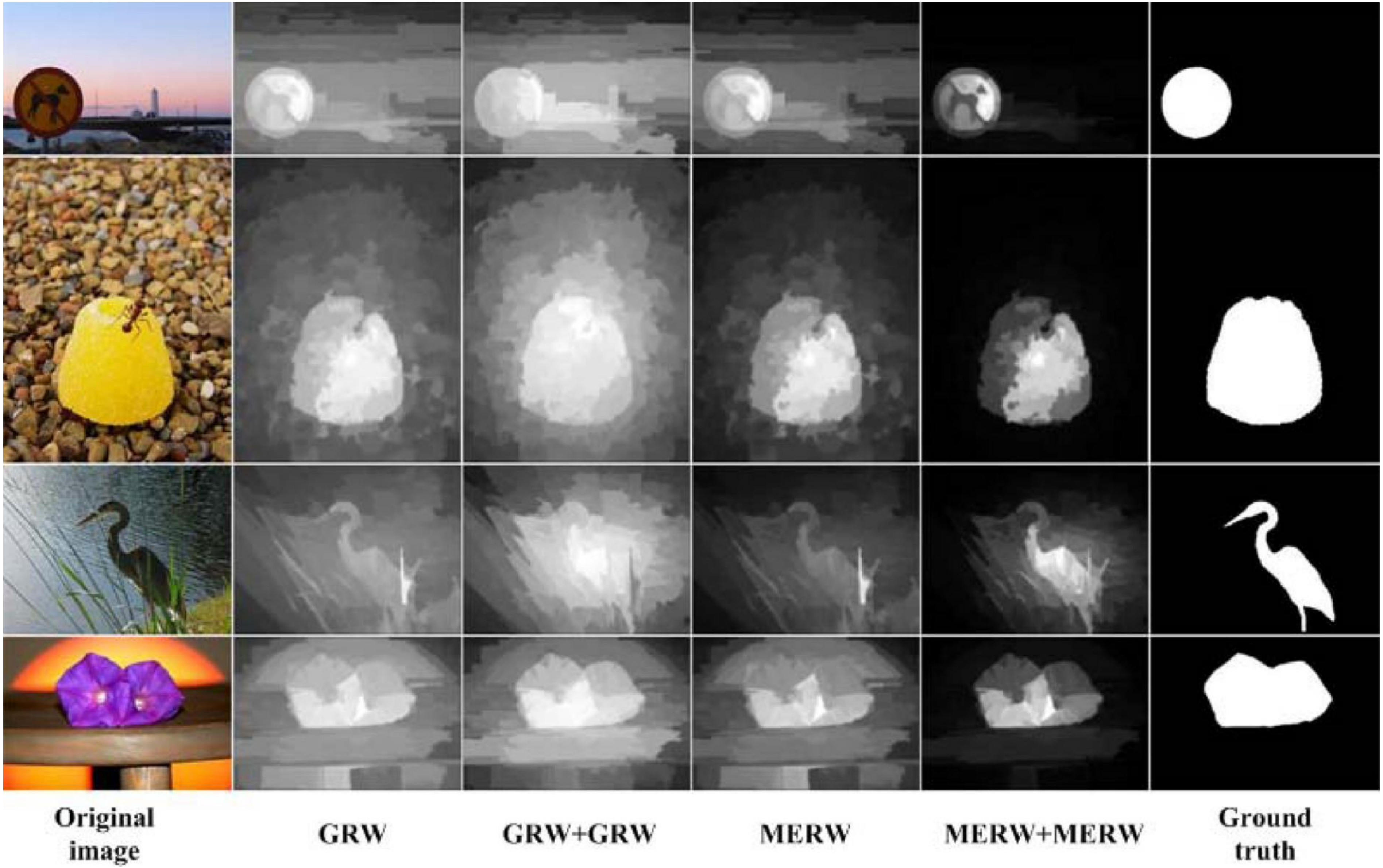
**The main "quantum corrections to stochastic models": localization, e.g. in semiconductor – where else it is essential?**

**some further work:**

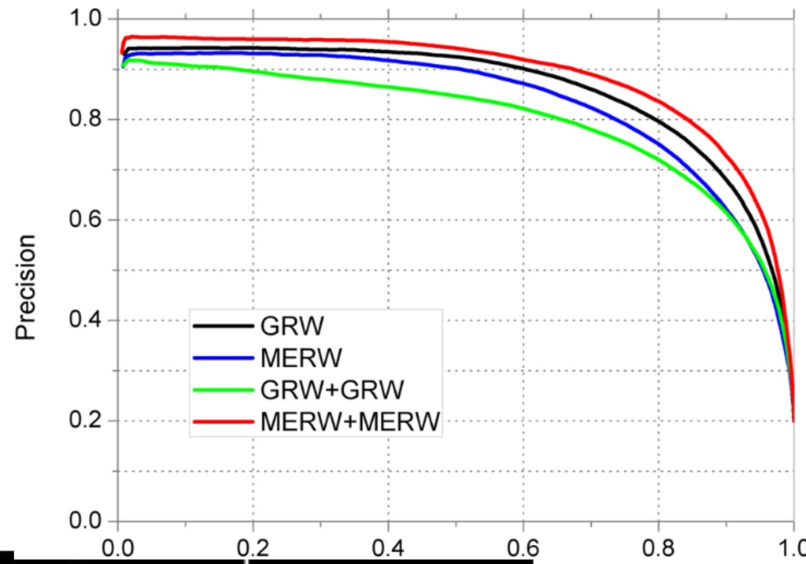
- improving mathematical formalism,
- try to motivate, derive Levy parameters from deeper dynamics,
- see $S^t(x,y) = \frac{\sum_i e^{-t\beta E_i} \langle x|\Psi_i\rangle\langle\Psi_i|y\rangle}{e^{-t\beta E_0}} \frac{\Psi_0(y)}{\Psi_0(x)}$ propagator as **"stochastic shift toward quantum eigenstate"** of perturbed trajectories,
- add velocity into consideration in analogy to Langevin equation,
- add other internal degrees of freedom like direction of spin,
- find deeper understanding of quantum mechanics,
- **find more quantum corrections to standard diffusion models**.

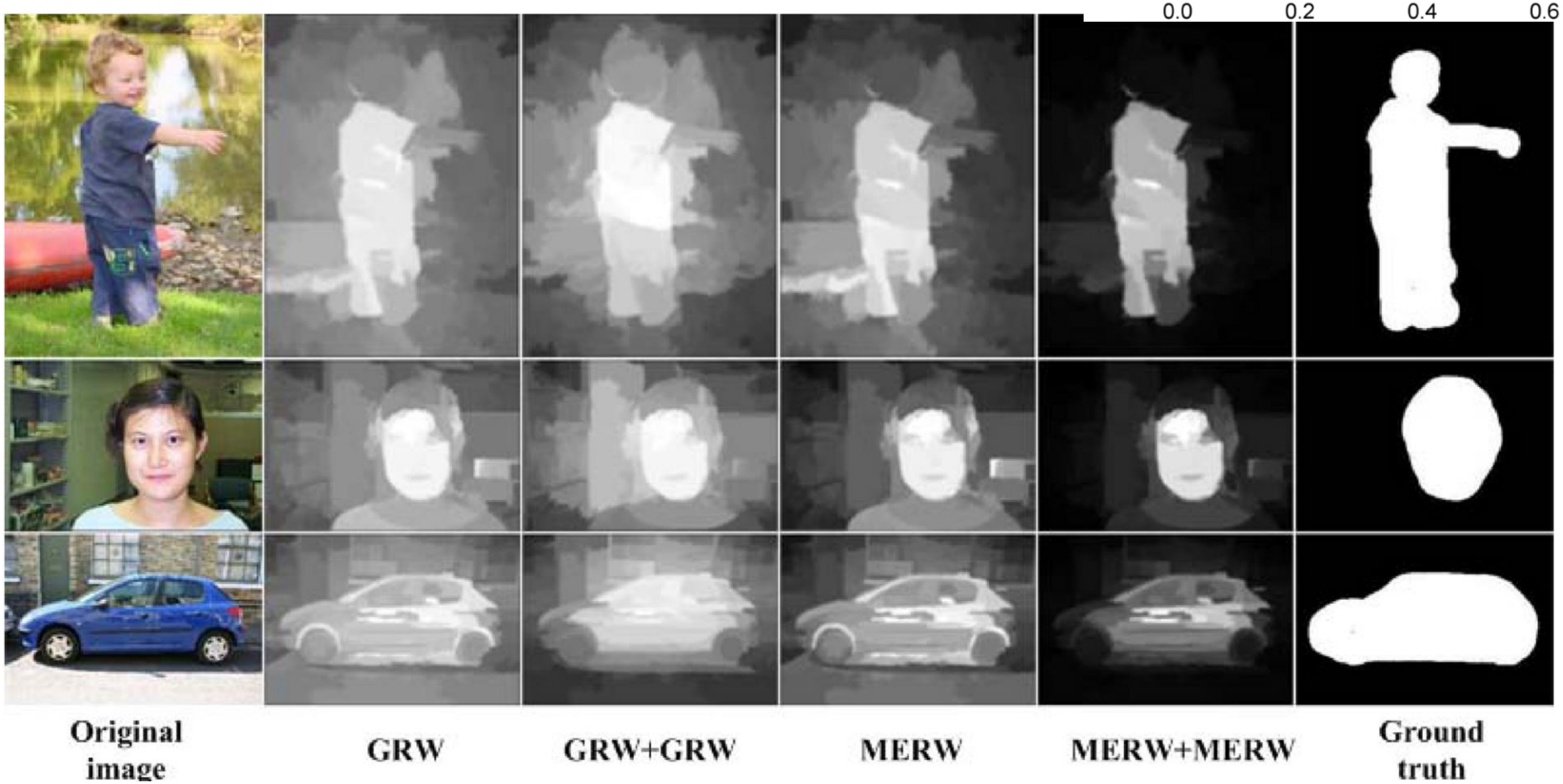
# Using MERW properties (localization) for various applications

JG Yu, J Zhao, J Tian, Y Tan, *Maximal Entropy Random Walk for Region-Based Visual Saliency* (IEEE, 2014)



- divide picture into regions (8x8 blocks, "superpixels")
- create graph among regions using similarities as weights ($w_{ij} = \exp(-d(r_i, r_j))$),
- saliency map is the stationary probability distribution of GRW or MERW

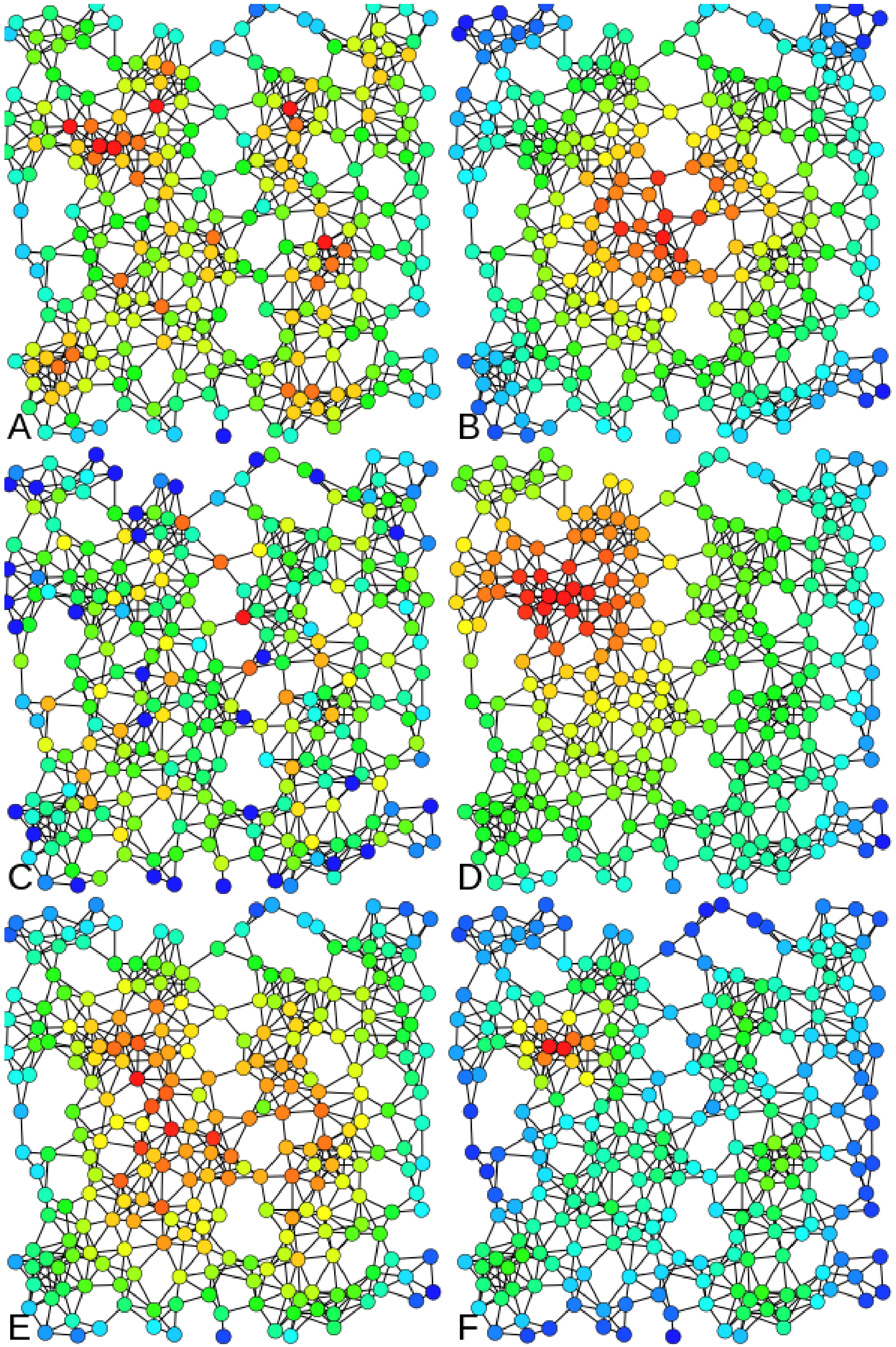


Centrality (graph theory, http://en.wikipedia.org/wiki/Centrality ):
**indicators which identify the most important vertices within a graph.**

Examples (for the same graph):
A) Degree centrality
($e.g. C(v) \propto \deg(v)$ – GRW),
B) Closeness centrality
($e.g. C(v) \propto \sum_{w\neq v} 1/d(v,w)$),
C) Betweenness centrality
(how many shortest paths go through $v$)
D) **Eigenvector centrality (MERW-like)**,
E) Katz centrality (e.g. PageRank),
F) Alpha centrality.

Drawing 2D diagrams for graphs:
positions from two high eigenvectors
(of $M$ or Laplacian: $L = \mathrm{diag}(\deg(i)) - M$ )



Delvenne, J.-C. & Libert, A.-S. *Centrality measures and thermodynamic formalism for complex networks,* Phys. Rev. E 83, 046117 (2011).

(e.g. Google) PageRank (GRW) → Entropy Rank (MERW)

($\alpha$ = Pr(going to a random page), $E = e^{-U_0}$ weight out of the graph edges)

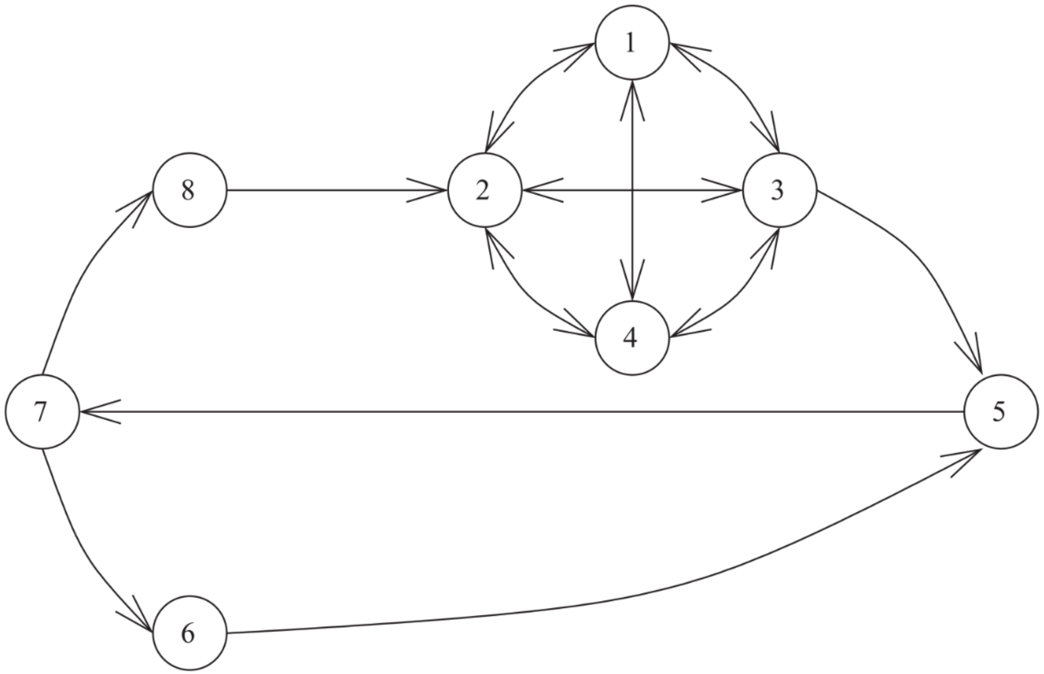
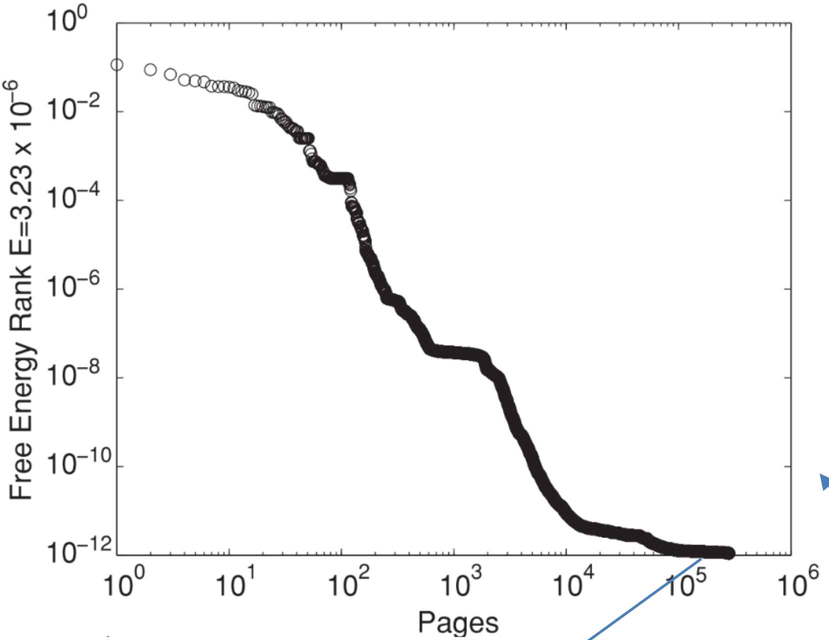



TABLE I. PageRank, free-energy rank, and entropy rank for the network of Fig. 1.

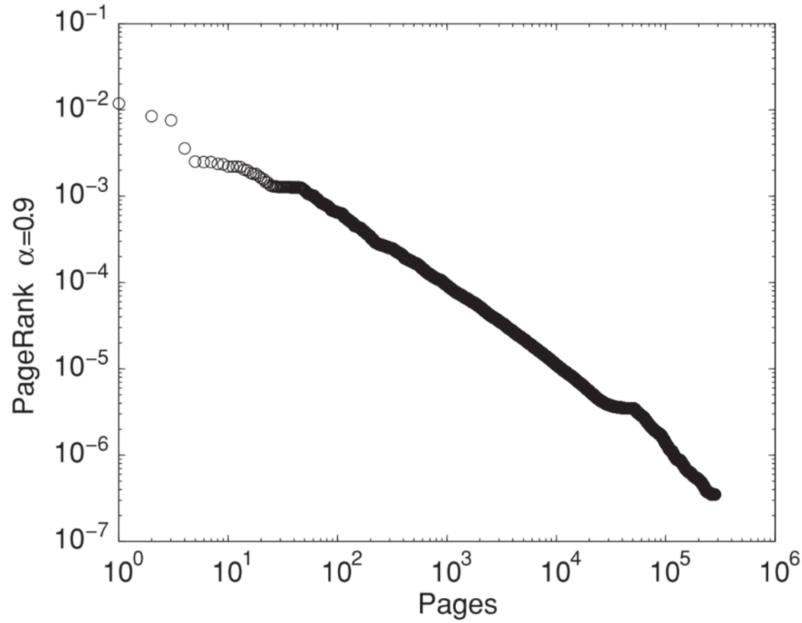
| Vertex | PageRank ($\alpha = 1$) | PageRank ($\alpha = 0.9$) | Entropy rank | Free-energy rank ($E = 0.03$) |
|---|---|---|---|---|
| 1 | 0.1705 | 0.1549 | 0.2464 | 0.2400 |
| 2 | 0.2045 | 0.1965 | 0.2487 | 0.2458 |
| 3 | 0.1818 | 0.1644 | 0.2487 | 0.2460 |
| 4 | 0.1705 | 0.1549 | 0.2464 | 0.2400 |
| 5 | 0.0909 | 0.1035 | 0.0032 | 0.0099 |
| 6 | 0.0455 | 0.0601 | 0.0001 | 0.0019 |
| 7 | 0.0909 | 0.1057 | 0.0032 | 0.0076 |
| 8 | 0.0455 | 0.0601 | 0.0031 | 0.0087 |

- vertex 8 becomes more interesting than 6 (pointing to “good pages”),
- cliques are swelling (localization) – problem with “link farms” …

**Experiments** on “289 000 – node piece of the Stanford web (http://www.kamvar.org/)”

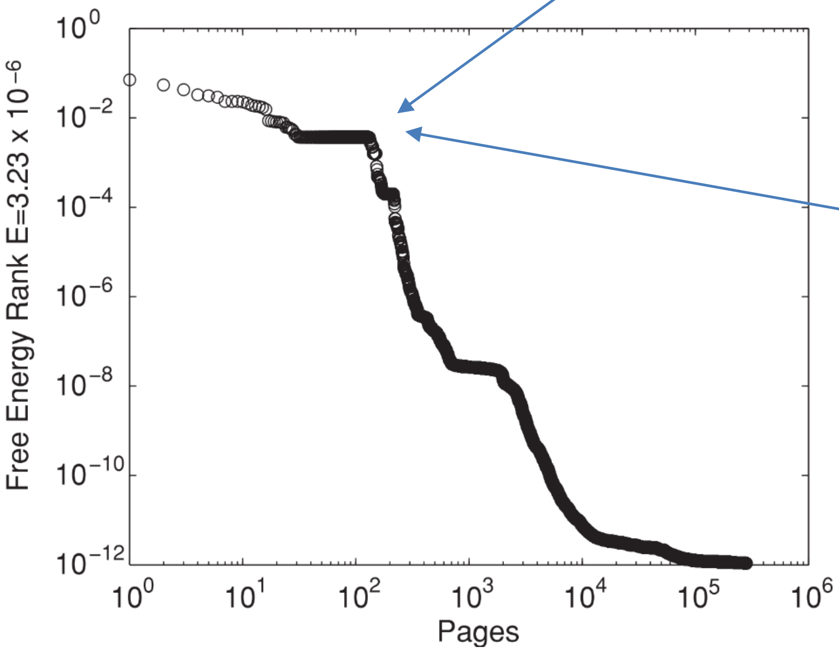



PageRank

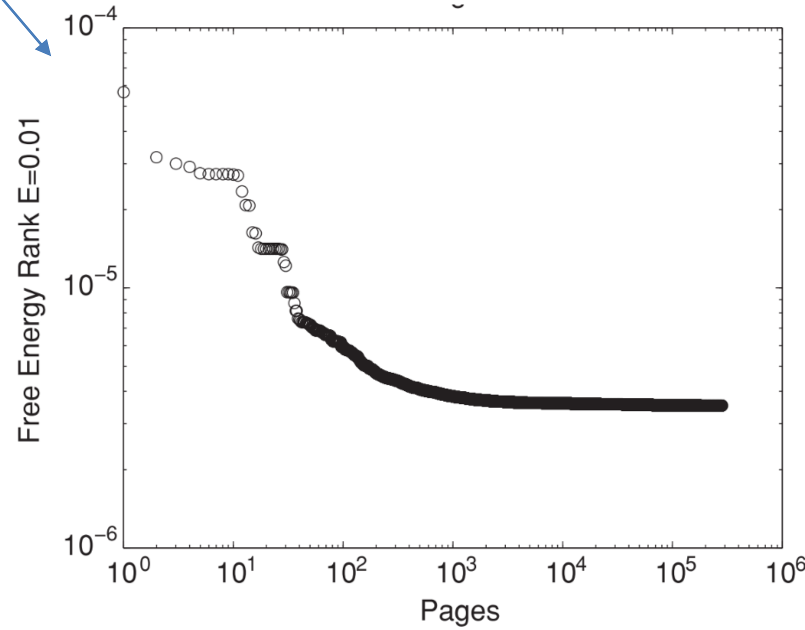



High E FER
(good for finding
best pages)

low H FER




low H FER
vertex with added
100 vert. clique
(“farm link”)
$200\,000th \to 627th$
(plateau → clique ?)



## **Mean first-passage time (MFPT)** (e.g. for community finding)

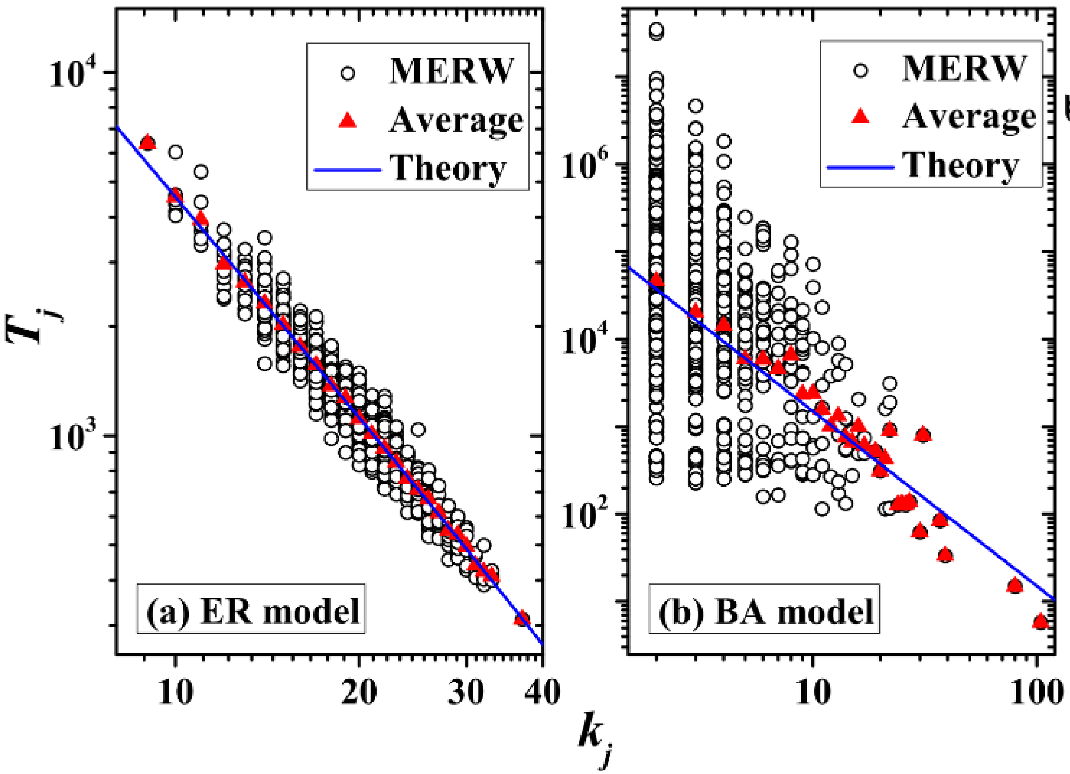
$M_{ij}$ – expected minimal time to reach vertex $j$ starting from $i$.

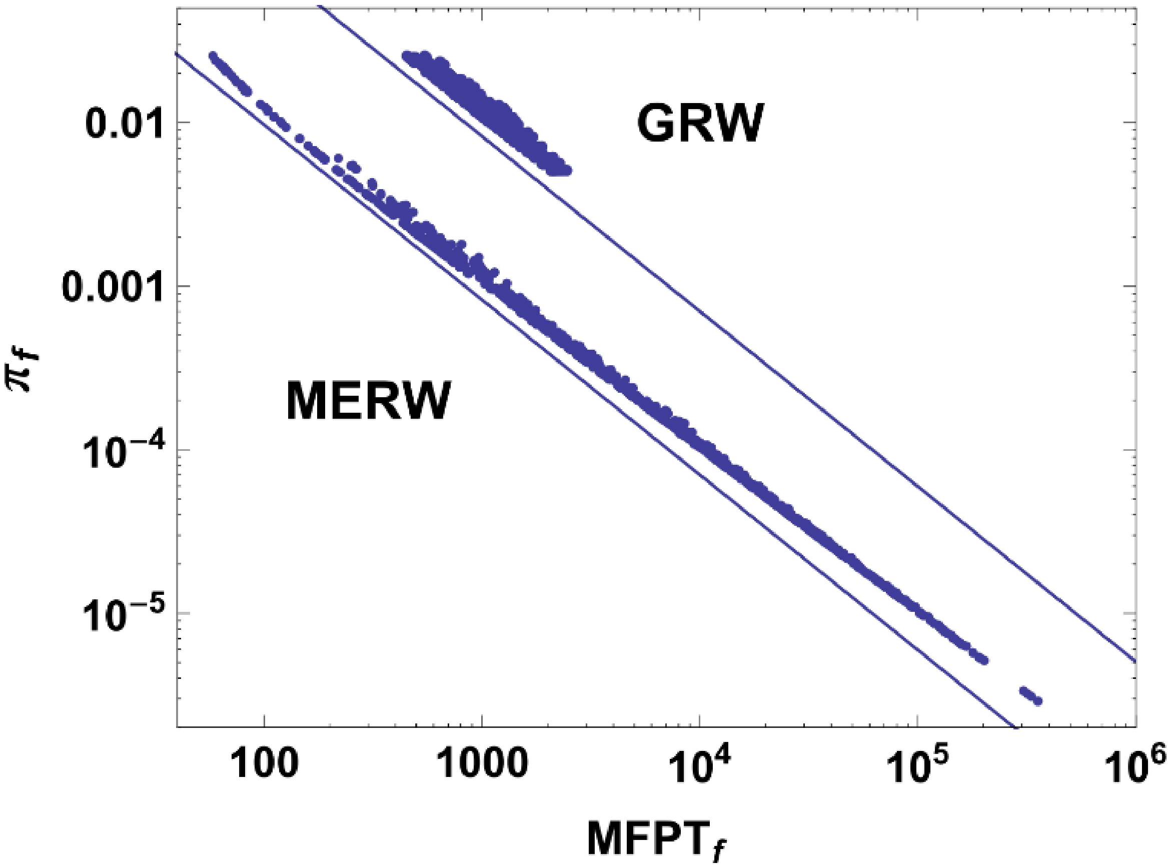
Y. Lin, Z. Zhang, *Mean first-passage time for maximal-entropy random walks in complex networks* (Nature, 2014)

Erdős–Rényi (ER): $\Pr(\to v_j) = \text{const}$

Barabási–Albert (BA): $\Pr(\to v_j) \propto k_j$

(scale-free : $P(k) \sim k^{-\gamma}$)




1000 vertices

J. Ochab, *Maximal-entropy random walk unifies centrality measures* (Phys. Rev. E, 2012)

**SimRank:** measure how similar two vertices are

G. Jeh and J. Widom. *Simrank: a measure of structural-context similarity* (KDD 2002)

$s(a,b) = \frac{C}{|N(a)||N(b)|} \sum_{x\in N(a)} \sum_{y\in N(b)} s(x,y)$ $\quad (1\ if\ a=b,\ \ 0\ if\ I(a)\cap I(b)=\emptyset)$

can be expressed by Expected$-f$ Meeting Distance (EMD) of two walkers $(a,b)$

$s'(a,b) = \sum_{t:(a,b)\rightsquigarrow(x,x)} P[t]\, f\big(l(t)\big)$ for $f(z)=z$ or $f(z)=C^z$

$P[t]$ - GRW probability of path $t$

**Link prediction** – which new interactions (links) are likely to occur?

The more similar they are, the more likely they will link

Li, R. H., Yu, J. X. & Liu, J. *Link prediction: the power of maximal entropy random walk* (ACM conference, 2011):

Replace GRW with MERW in $P[t]$, getting $S(a,b) = \frac{C\psi_a\psi_b}{\lambda^2} \sum_{x\in N(a)} \sum_{y\in N(b)} \frac{S(x,y)}{\psi_x\psi_y}$

Uniform probability distribution among paths (MERW) instead of edges

# 27 link prediction methods (the higher the better), “ME” – maximal entropy

| SM | ER | BA | SW | USAir | C.ele | Yeast | Power | NetSci | GrQc | HepPh | HepTh |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CTT | 0.710 | 0.750 | 0.791 | 0.847 | 0.784 | 0.709 | 0.713 | 0.917 | 0.520 | 0.523 | 0.525 |
| **CTME** | 0.720 | 0.746 | 0.745 | 0.855 | 0.798 | 0.501 | 0.501 | 0.866 | 0.556 | 0.645 | 0.534 |
| CK | 0.805 | 0.883 | 0.804 | 0.856 | 0.809 | 0.715 | 0.501 | 0.799 | 0.513 | 0.501 | 0.513 |
| **MECK** | 0.940 | 0.981 | 0.845 | 0.936 | 0.856 | 0.757 | 0.501 | 0.975 | 0.517 | 0.501 | 0.503 |
| NCK | 0.502 | 0.501 | 0.501 | 0.708 | 0.706 | 0.501 | 0.501 | 0.501 | 0.503 | 0.508 | 0.501 |
| **NMECK** | 0.903 | 0.983 | 0.982 | 0.931 | 0.969 | 0.710 | 0.501 | 0.971 | 0.623 | 0.750 | 0.675 |
| DK | 0.835 | 0.813 | 0.983 | 0.836 | 0.838 | 0.829 | 0.764 | 0.965 | 0.501 | 0.605 | 0.593 |
| **MEDK** | 0.999 | 0.983 | 0.998 | 0.991 | 0.971 | 0.749 | 0.812 | 0.963 | 0.739 | 0.735 | 0.746 |
| NDK | 0.786 | 0.711 | 0.956 | 0.920 | 0.778 | 0.731 | 0.857 | 0.908 | 0.531 | 0.530 | 0.530 |
| **NMEDK** | **0.999** | **0.983** | **0.998** | **0.997** | **0.978** | **0.970** | 0.857 | **0.996** | **0.739** | 0.755 | **0.758** |
| RK | 0.851 | 0.907 | 0.973 | 0.898 | 0.887 | 0.803 | 0.864 | 0.624 | 0.632 | 0.608 | 0.561 |
| **MERK** | 0.999 | 0.983 | 0.998 | 0.981 | 0.949 | 0.812 | 0.812 | 0.963 | 0.618 | 0.745 | 0.735 |
| NRK | 0.504 | 0.501 | 0.501 | 0.719 | 0.501 | 0.703 | 0.806 | 0.501 | 0.501 | 0.508 | 0.504 |
| **NMERK** | 0.999 | 0.983 | 0.998 | 0.983 | 0.975 | 0.968 | 0.857 | 0.986 | 0.739 | 0.755 | 0.756 |
| MENK | 0.999 | 0.983 | 0.998 | 0.936 | 0.975 | 0.799 | 0.812 | 0.963 | 0.618 | 0.730 | 0.746 |
| NNK | 0.503 | 0.501 | 0.501 | 0.819 | 0.501 | 0.705 | 0.806 | 0.501 | 0.501 | 0.508 | 0.504 |
| **NMENK** | 0.999 | 0.983 | 0.998 | 0.983 | 0.965 | 0.965 | 0.857 | 0.996 | 0.739 | 0.755 | 0.752 |
| PD | 0.926 | 0.974 | 0.953 | 0.971 | 0.866 | 0.887 | 0.857 | 0.722 | 0.666 | 0.618 | 0.628 |
| **MEPD** | 0.999 | 0.976 | 0.998 | 0.993 | 0.964 | 0.968 | 0.857 | 0.913 | 0.739 | 0.755 | 0.758 |
| PDM | 0.805 | 0.764 | 0.957 | 0.972 | 0.798 | 0.886 | 0.857 | 0.874 | 0.616 | 0.660 | 0.530 |
| **MEPDM** | 0.999 | 0.983 | 0.998 | 0.990 | 0.976 | 0.970 | 0.857 | 0.996 | 0.739 | 0.755 | 0.758 |
| SR | – | – | – | 0.905 | 0.860 | – | – | 0.955 | – | – | – |
| **MESR** | – | – | – | 0.960 | 0.876 | – | – | 0.963 | – | – | – |
| CN | 0.884 | 0.782 | 0.501 | 0.386 | 0.971 | 0.752 | 0.802 | 0.961 | 0.617 | 0.623 | 0.635 |
| AA | 0.886 | 0.781 | 0.501 | 0.409 | 0.975 | 0.793 | 0.806 | 0.969 | 0.623 | 0.630 | 0.638 |
| HPLP+ | 0.983 | 0.971 | 0.978 | 0.979 | 0.974 | 0.965 | **0.886** | 0.984 | 0.725 | 0.753 | 0.732 |
| SRW | 0.991 | 0.977 | 0.989 | 0.983 | 0.972 | 0.967 | 0.863 | 0.983 | 0.731 | **0.760** | 0.754 |

# MERW – the most random among random walks

uniform distribution among paths, not edges (GRW)

- As the choice of statistical parameters of an **informational channel**

**MERW** allows to maximize channel capacity under some constraints (language?)

- **As random walk/diffusion** (scale-free)

**GRW**: the walker indeed performs succeeding random decisions

**MERW**: only represents our (lack of) knowledge about a complex dynamics

- **For metrics to analyze complex network**

**GRW** sees only degrees of vertices

**MERW** allows to evaluate importance in the space of possible paths

- social/evolutionary entropy (Lloyd Demetrius):

"thinking" in terms of paths (reason→result chains) of possibilities?

**GRW → MERW**

**in many cases improves performance or agreement**