”Project 42”

Heikki Hyötyniemi
TKK / Control Engineering

Presentation at Control Seminar
September 22, 2006

HELSINKI UNIVERSITY OF TECHNOLOGY
Control Engineering Laboratory
Cybernetics Group

- As you will see, Control Engineering Laboratory is very successful in mastering projects and education
- But who came to University thinking it is something *more* ...?
- ... Indeed, there are many hidden structures and networks

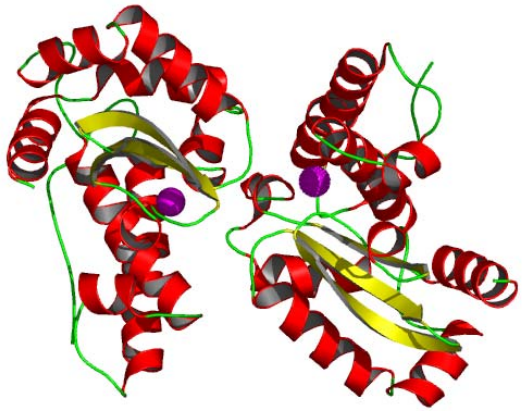
- Here, some *undercurrents* at the Lab are presented

Webs of wisdom

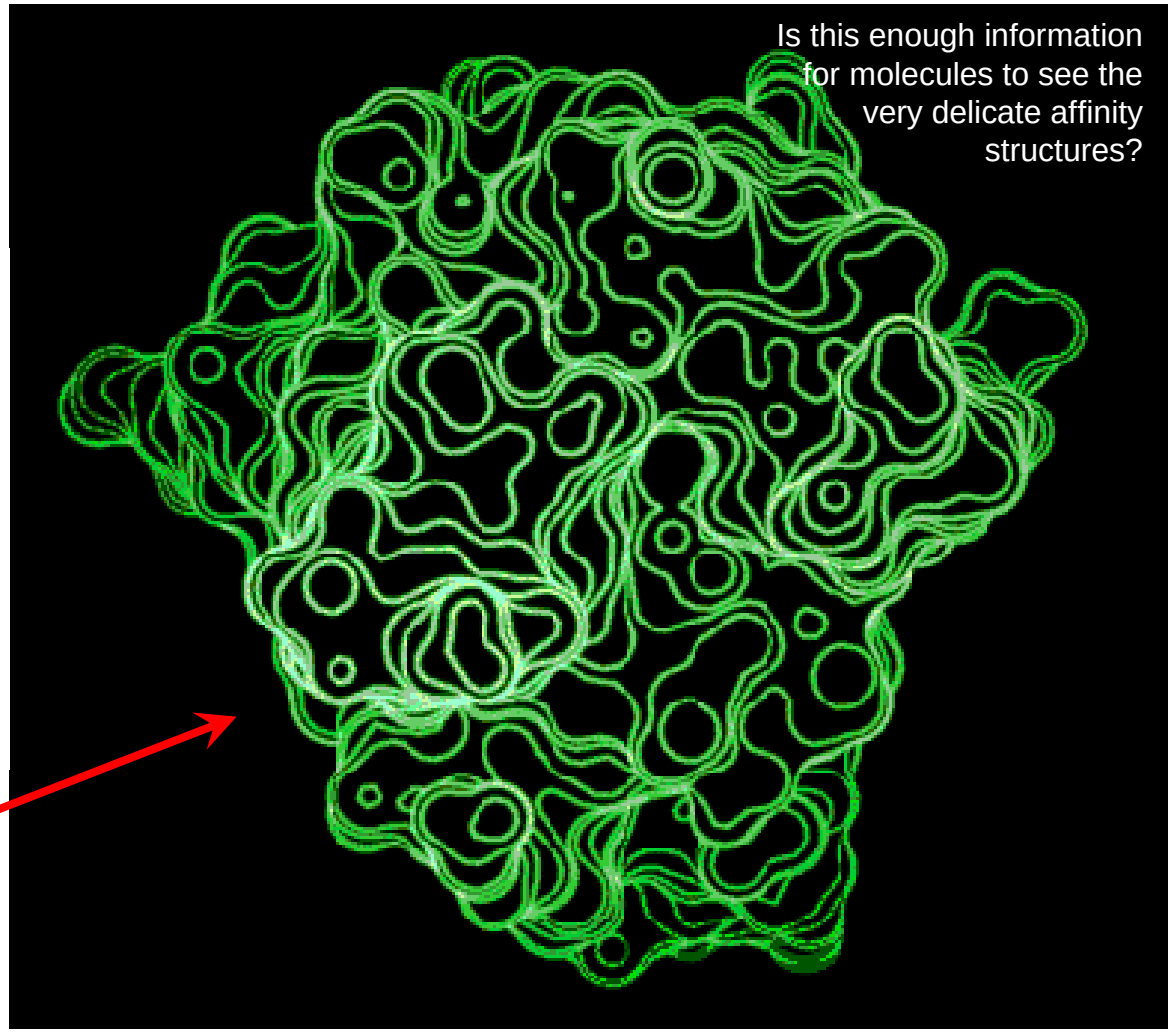
- Aristotle : ”Heart is the home of soul”
  - Heart is in the ”innermost” organ
  - Speech comes from the chest, where the heart is
  - Heartbeat accelerates when one is excited, etc.
  - Brain is only needed for cooling of blood!
- Aristotle was the big authority for more than 1000 years, offering the most logical explanations at that time
  - Before gravitation law, based on the Aristotelian world view, the best explanations based on flat Earth hypothesis (objects want to fall ”down”)
- Further: Before the theory of relativity, the best explanation for diversity of species *was* divine (there is not enough coal in the Sun to last for millions of years)

- One's thinking is bound to one's own world view; are we now on the correct track?
- Thinking patterns 500 years ago seem so ridiculous – what do they think about *us* 500 years from now in the future?
- Today there are so many new incompatible observations that one can say that there are more mysteries than ever before
- The "best explanations" are probably to be changed again
- Evidence & explanations are not yet in balance – examples:
  - *Gene transcription* + *translation* – intelligence needed in coordination!?
  - *Proteins* + *enzymes* – huge number of functionalities: Pattern recognition?!
  - How to understand and model *protein folding*?
  - What is the nature of *orbitals*, the predestinated structures in molecules?

- For example: Enzyme *superoxide dismutase*



- Only electric fields can be experienced by other molecules



”Pallas Athene Hypothesis”

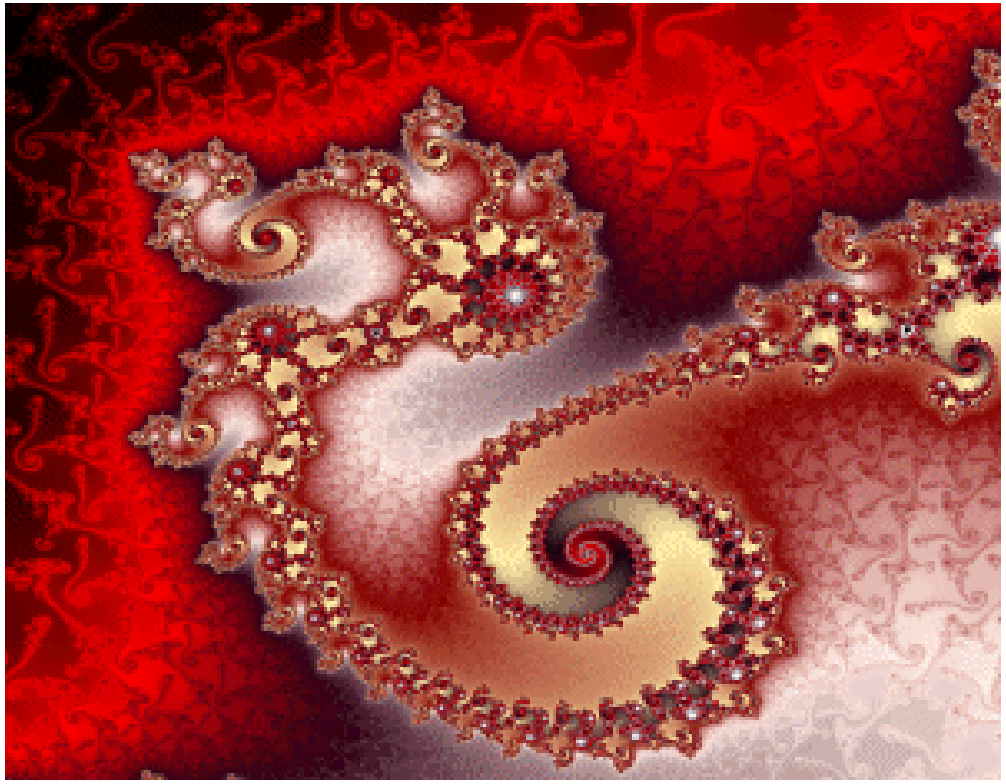
- Today, complex phenomena can be described but they cannot be really modeled

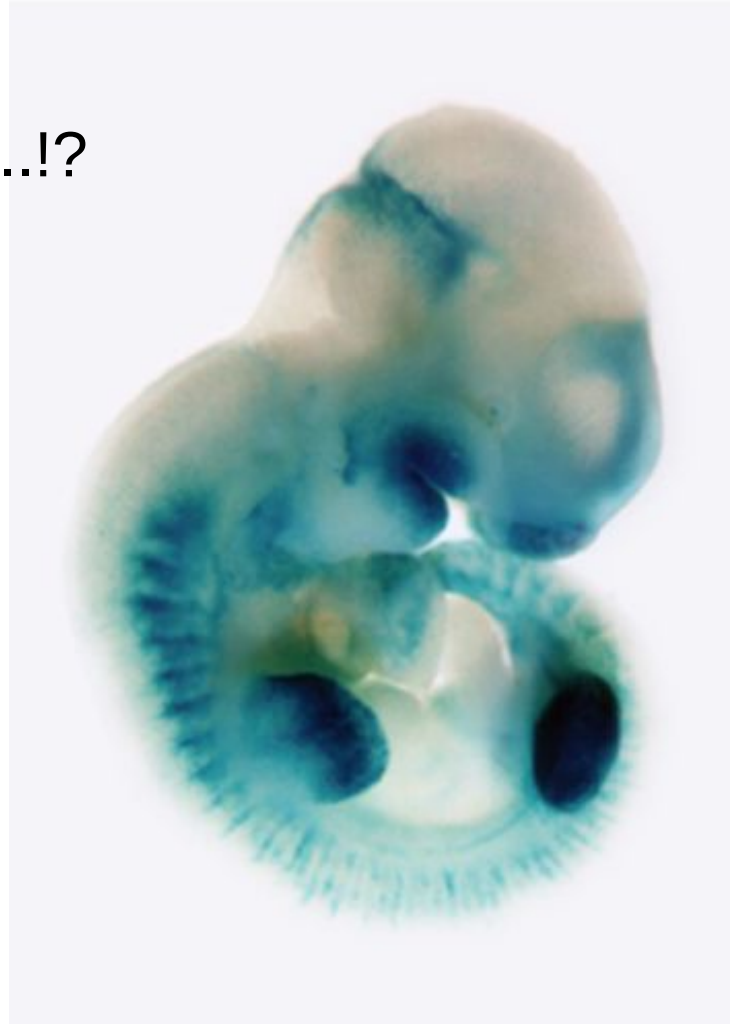
Assume there exists a general theory of complex systems

Further, assume that there exists mathematics for analysis and synthesis of such systems

Intuitions that collapse

- There must be something in common beyond complex systems ...!?





Surface forms similar – functions very different

More appropriate starting points

- Basic mystery: How can the global-level expressions be implemented by the local-level actors with no global control?

- The local actors can only react to local gradients – the system is characterized by (generalized) diffusion processes

- Observed behaviors are result of balance of tensions among the system and its environment

- Interpret static equations as dynamic equilibria: Emergent patterns reflect underlying dynamic attractors

From static pattern to a dynamic one

- Assume the system reacts (linearly) to its environment:

$$\bar{x} = \phi^T u$$

Standard way to characterize a system

- Assume that the system is restructured appropriately:

$$A\bar{x} = Bu$$

Tension equilibrium

- Assume that the balance is not yet reached:

$$\frac{dx}{\gamma\, dt} = -Ax + Bu$$

Diffusion process

- For such gradient, there is a *cost* characterizing the system:

$$J = \frac{1}{2} x^T A x - x^T B u$$

Opposite way to characterize a system!

How to interpret

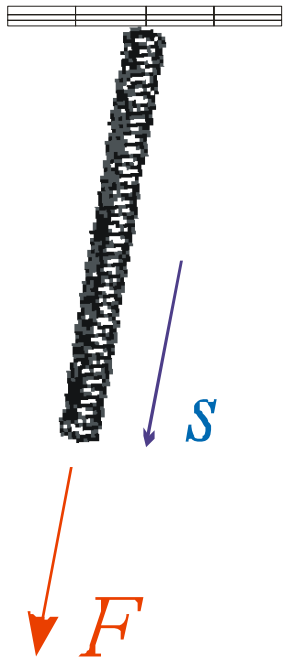
- Study a one-dimensional case: Spring (spring constant k) stretched (deformation s) by an external force F
- There are *external* and *internal* stored energies in spring (zero level = zero force):

1. Due to the external potential field

$$W_{\text{ext}} = -\int_0^s F\, ds = -Fs$$

2. Due to the internal tensions

$$W_{\text{int}} = \int_0^s ks\, ds = \frac{1}{2} ks^2$$



- Generalization: There are many forces, and many points
- Spring between points s_1 and s_2 (can also be torsional, etc.)

$$W_{\text{int}}(s_1, s_2) = \frac{1}{2} k_{1,2} \left(s_1 - s_2 \right)^2 = \frac{1}{2} k_{1,2} s_1^2 - k_{1,2} s_1 s_2 + \frac{1}{2} k_{1,2} s_1^2$$

- A matrix formulation is also possible:

$$W_{\text{int}}(s) = \frac{1}{2} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}^T A \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} \qquad W_{\text{ext}}(s, F) = - \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}^T B \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix}$$

- F_j: Virtual ”generalized forces” as projected along the directions of movements – also torques, shear stresses, etc., all presented in the same framework (for linear structures)

”All” complex systems are elastic systems!

- Now: The difference of potential energies can be expressed as

$$J(s,F) = \frac{1}{2} s^T A s - s^T B F$$

The same cost as found above!

- Here, A is *matrix of elasticity*, and B determines projections
- Matrix A must be symmetric, and must be positive definite to represent stable structures sustaining external stresses
- Principle of minimum potential (deformation) energy: *Structure under pressure ends in minimum of this criterion*
- Elastic systems yield when pressed, but bounce back after it
- Are there additional intuitions available?

Assumption: Goals of local scale actors

- Compare to gravitational field: Potential energy is

  $W_{\text{pot}} = mg\ \Delta h$ *"force times deformation"*

- Elastic system: Average transferred energy / power

  $\mathrm{E}\{x_i u_j\}$

- Now assume:

  System tries to maximize the coupling with its environment

- That is:

  Maximize the average product of action and reaction

- If this holds for all actors, the system matrices can be written $A = \beta\,\mathrm{E}\{xx^T\}$ and $B = \beta\,\mathrm{E}\{xu^T\}$ for some scalar β

Towards abstraction level #2

- Cybernetic model = statistical model of balances of $x(u)$
- Assume dynamics of u is essentially slower than that of x and study the covariance of $x = \phi^T u = \mathrm{E}\{xx^T\}^{-1}\,\mathrm{E}\{xu^T\}\,u$

$$\mathrm{E}\{xx^T\} = \mathrm{E}\{xx^T\}^{-1}\,\mathrm{E}\{xu^T\}\,\mathrm{E}\{uu^T\}\,\mathrm{E}\{xu^T\}^T\,\mathrm{E}\{xx^T\}^{-1}$$

or

$$\mathrm{E}\{xx^T\}^3 = \mathrm{E}\{xu^T\}\,\mathrm{E}\{uu^T\}\,\mathrm{E}\{xu^T\}^T$$

or

$$\left(\phi^T\,\mathrm{E}\{uu^T\}\,\phi\right)^3 = \phi^T\,\mathrm{E}\{uu^T\}^3\,\phi \qquad n < m$$

- Balance on the statistical level = *second-order balance*

Solution

- Expression fulfilled for $\phi = \theta_n D$, where θ_n is a matrix of n of the covariance matrix eigenvectors, and D is orthogonal
- This is because left-hand side is then

$$\left(\phi^T \mathrm{E}\left\{uu^T\right\}\phi\right)^3 = \left(D^T\theta_n^T \mathrm{E}\left\{uu^T\right\}\theta_n D\right)^3 = \left(D^T \Lambda_n D\right)^3 = D^T \Lambda_n^3 D$$

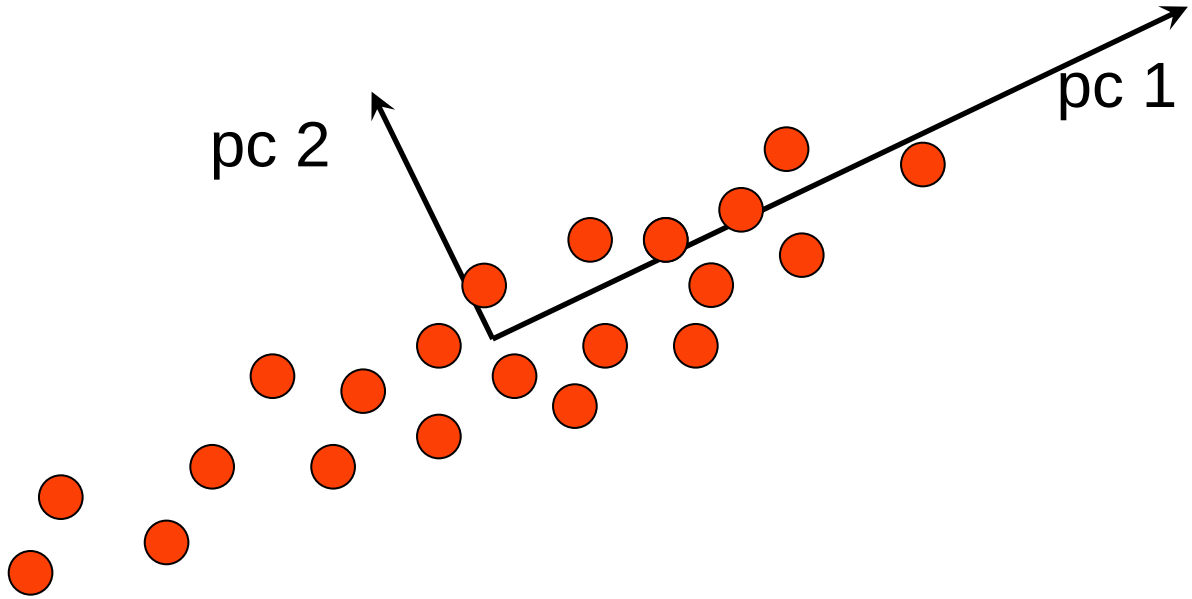
- and right-hand side is

$$\phi^T \mathrm{E}\left\{uu^T\right\}^3 \phi = D^T\theta_n^T \mathrm{E}\left\{uu^T\right\}^3 \theta_n D = D^T \Lambda_n^3 D$$

- Stable solution when θ_n contains the *most significant* data covariance matrix eigenvectors

Principal components

- Principal Component Analysis = Data is projected onto the most significant eigenvectors of the data covariance matrix
- This projection captures maximum of the variation in data
- Principal subspace = PCA basis vectors rotated somehow



Example case: Hebbian learning

- The Hebbian learning rule (by physician Donald O. Hebb) dates back to mid-1900's:

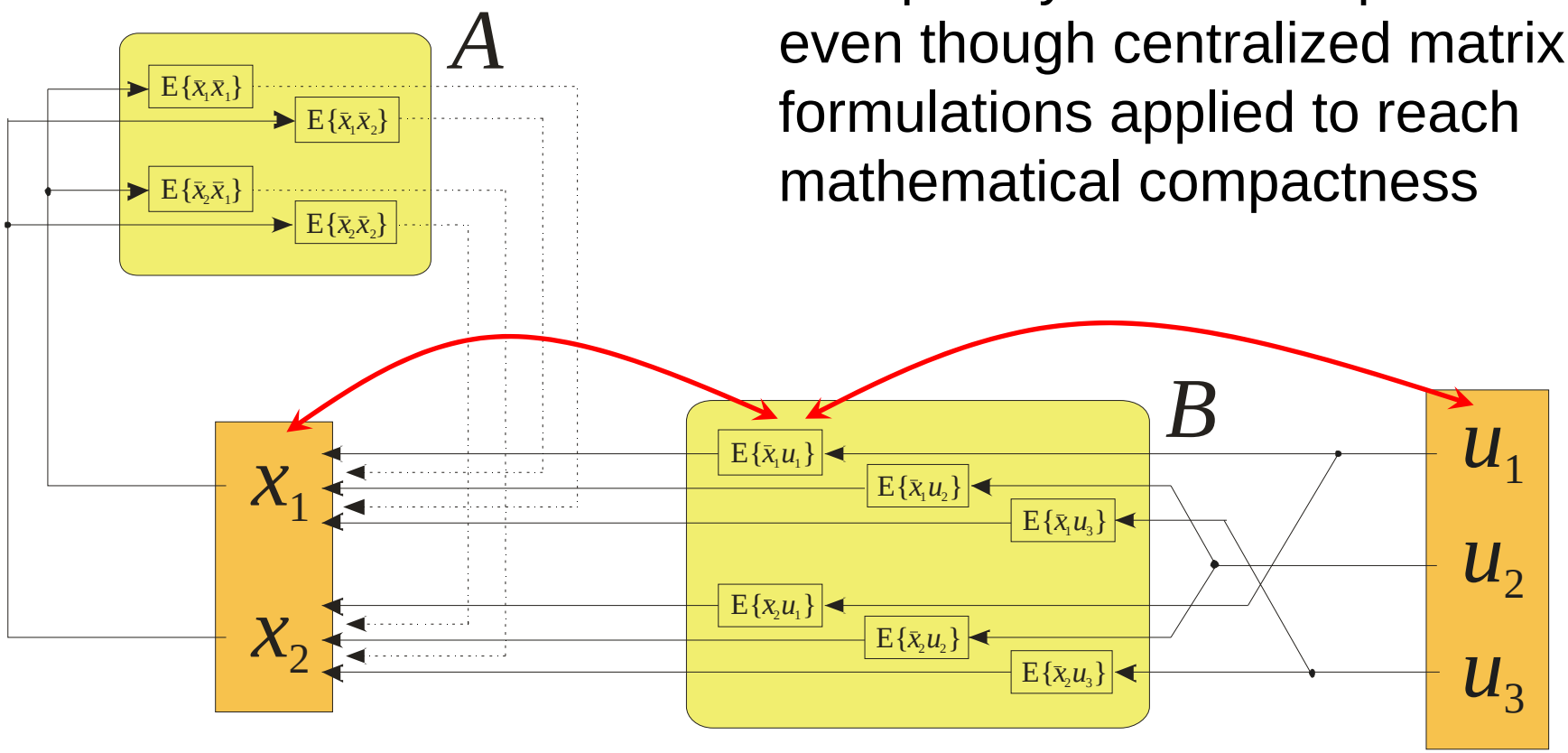
  "If the neuron activity correlates with the input signal, the corresponding synaptic weight increases"

- PCA based modeling of input data takes place in the brain?
- Powerful intuitions available concerning other cybernetic systems as well: Construction of the PCA model means best possible exploitation of resources and evolutionary benefits

Hebbian/anti-Hebbian system

$$\dot{x} = -\mathrm{E}\left\{xx^T\right\}x + \mathrm{E}\left\{xu^T\right\}u$$

- Explicit feedback structures
- Completely localized operation, even though centralized matrix formulations applied to reach mathematical compactness



Extension to other domains

- Theodosius Dobzhansky: ”Nothing in biology makes sense without reference to evolution”
- Extension: ”Nothing in complex systems makes sense without reference to evolution”

- It can be claimed that evolutionarily surviving systems implement the derived framework
- Employing the presented model framework, there is best possible exploitation of resources

- Completely local operation: ”Go towards resources, avoid competition”

Properties of the model

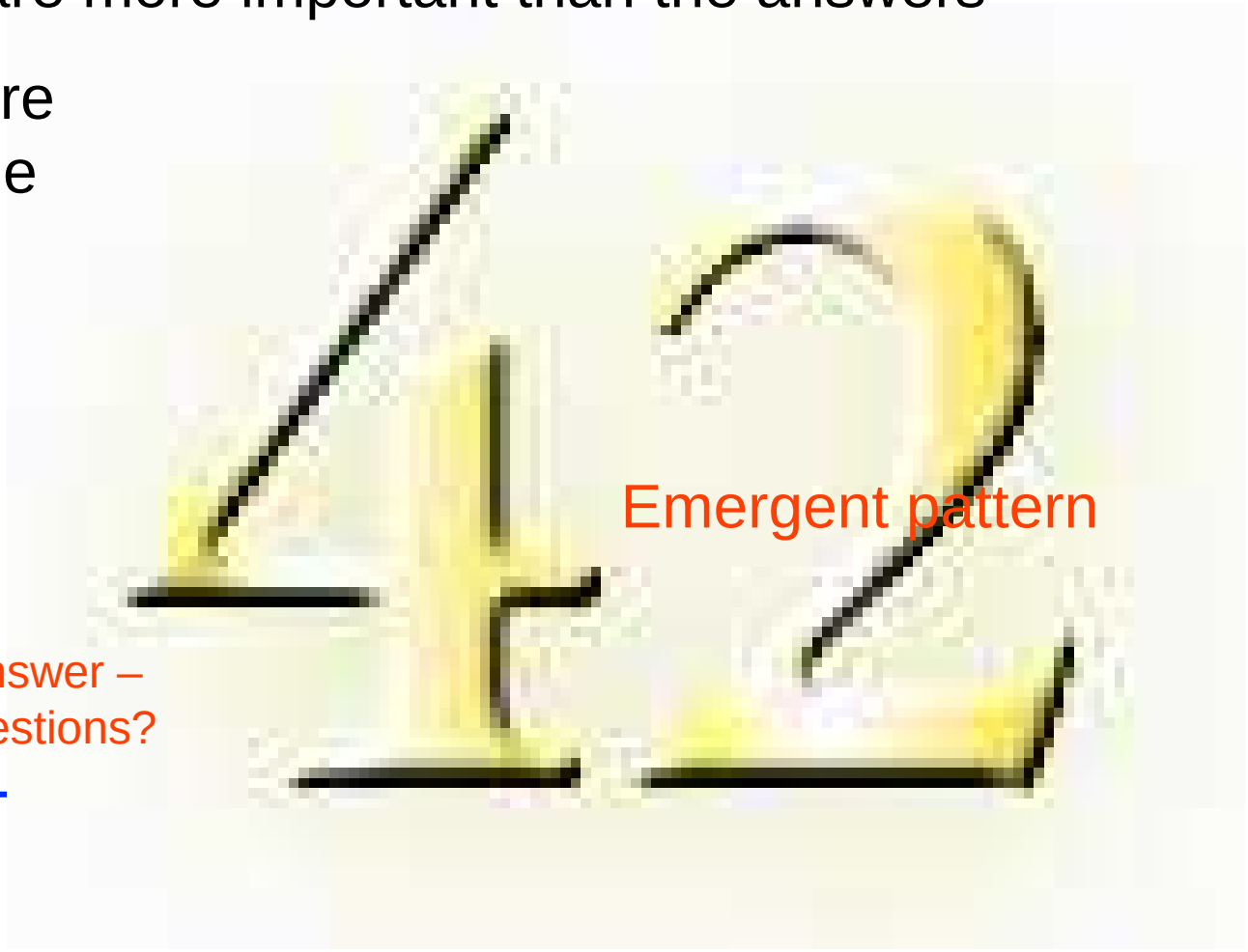
- Robustness.
  - In nature, no catastrophic effects typically take place; even key species are substituted if they become extinct (after a somewhat turbulent period)
  - Now, this can also be explained in terms of the principal subspace: If the profiles are almost orthogonal (PCA-like), disturbances do not cumulate
  - Also because of the principal subspace, sensitivity towards random variations are suppressed
- Biodiversity.
  - In nature, there are many competing species, none of them becoming extinct; modeling this phenomenon seems to be extremely difficult
  - Now, this results from the principal subspace nature of the model: As long as there are various degrees of freedom in input, there are different populations
  - Within populations, this also explains why there exists variation within populations as the lesser principal components also exist ...

Elastic systems

- New interpretation of cybernetic systems –
- **”First-order cybernetic system”**
  - Finds balance under external pressures, pressures being compensated by internal tensions
  - Any existing (complex) interacting system that maintains its integrity!
  - Implements minimum *observed* deformation energy
- **”Second-order cybernetic system”**
  - Adapts the internal structures to better match the observed environmental pressures – towards *maximum experienced stiffness*
  - Any existing (competing) interacting system that has survived in evolution!
  - Implements minimum *average* observed deformation energy

”Life, Universe, and Everything”

- At least some of the universal problems of complex systems can be addressed in the framework of neocybernetics
- Good questions are more important than the answers
- Such questions are searched for in the project



We already know the answer –
what are the correct questions?

About scientific discovery

- What is needed for scientific work?

1. Drive
   - Getting acquainted with very different things – new knowledge gives new ”eye-glasses” to see the world through
   - Curiosity + eternal inspiration
   - Perspiration: Most ideas are no good – one needs stamina to continue, wisdom to give up!
2. Direction

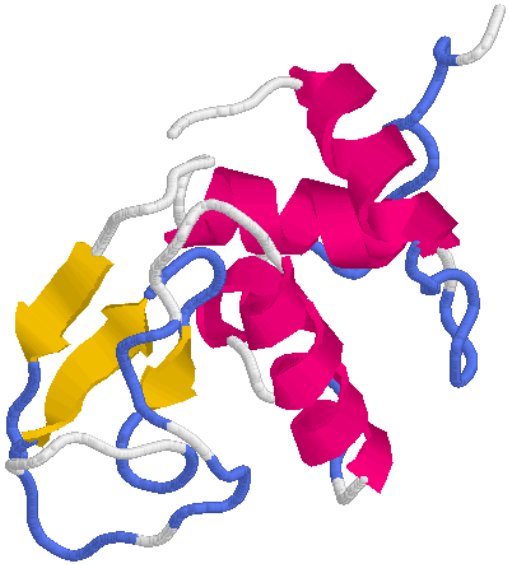
– Giving guidelines is an intellectual contradiction!

   - Where nobody has gone before / what nobody has thought of before!
   - Aesthetics: See the ”big picture” – and find connection between patterns:

$$J = x^T \mathrm{E}\{xx^T\}x - x^T \mathrm{E}\{xu^T\}u$$

Example: Analysis of *orbitals*

- Are orbitals predetermined structures hosting electrons?
- Or are they just emergent phenomena reflecting more fundamental underlying processes?
- Study what kind of consequences it has if a molecule is regarded as a (truly) cybernetic population of electrons

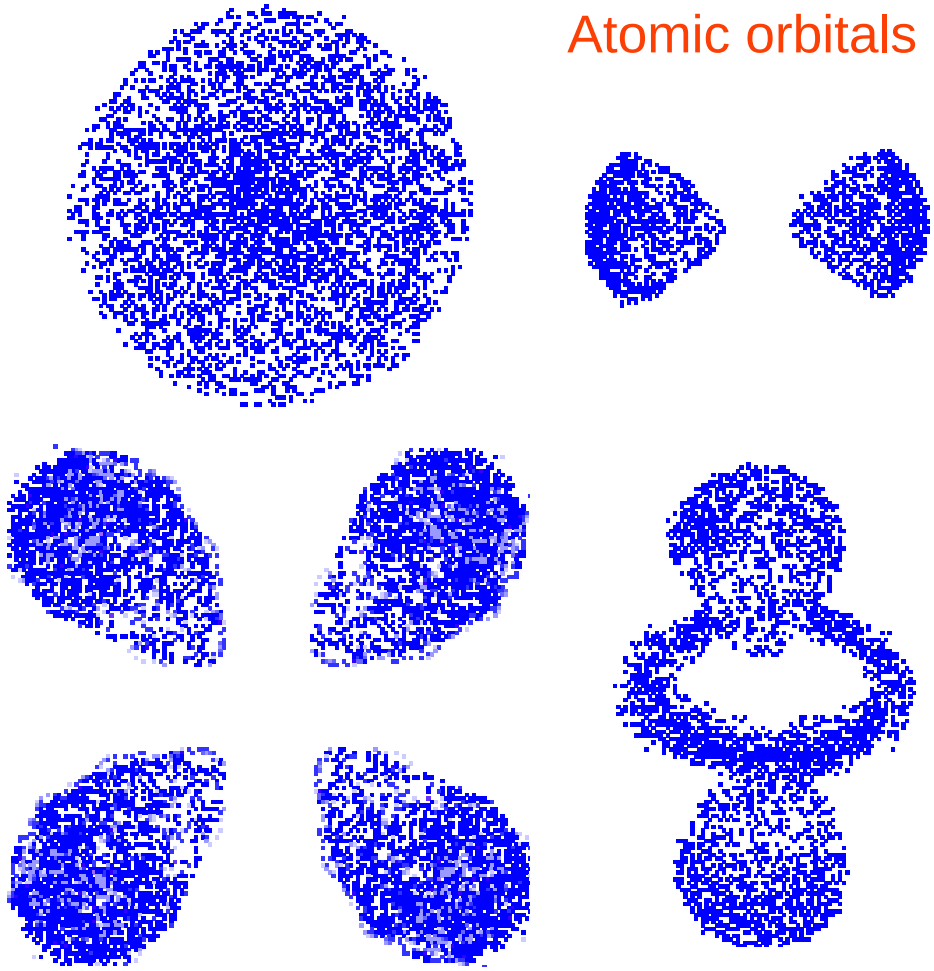


Applications:
Modeling the protein folding?
Understanding catalysis?

- Electrons are *delocalized* around nuclei
- Orbitals = ”probability distributions of electrons”
- Molecular orbitals = sums of atomic orbitals?

BUT:

- Molecular level is yet another emergent level
- Distributions extend over the whole molecule



- The molecular orbitals cannot directly (or most efficiently) be studied in terms of atom orbitals: Strange ”hybridisations”, etc., need to take place ...
- Assume that the quantum phenomena also *can be* modeled efficiently
- Assume it is simply a play among independent local-scale electric fields that is taking place in a molecule
- Then it helps when there is a strong structural framework as a target = neocybernetic model
- The model structure dictates the ways to interpret behaviors – an interesting question is whether these interpretations can be approved

Macroscopic analysis of electric fields

- Assume that there are various overlapping electric fields, and let $x_i(t)$ denote the electric charge within the field i.
- Energy that is stored in the potential fields:

  1. Within a single charge field

  $$J_{i,i} = c \int_0^{x_i} \xi \, d\xi = \tfrac{1}{2} c \, x_i^2$$

  2. Among overlapping fields

  $$J_{i,j} = c \int_0^{x_i} x_j \, d\xi = c \, x_i x_j$$

- If charges of i and j have the same sign, potential is positive, denoting repulsion; otherwise there is attraction

Microscopic analysis

- However, in microscopic scale, there are no charges to be observed, only interactions
- Now let $x_i(t)$ denote the momentary field strength within the field ("orbital") number i
- Macroscopic phenomena = long-term averages over time axis
- Assume that $p_{i,j}$ is the overall interaction probability among orbitals i and j
- Total energy that is stored in the potential fields can be expressed as

$$J' = p_{1,1}J_{1,1} + p_{1,2}J_{1,2} + \cdots + p_{n,n}J_{n,n}$$

- Because of the dual interpretation of the orbitals (charge distribution *and* probability distribution), one can express the joint distribution, or long-term mutual interaction (assuming independence) as (α being some constant)

$$p_{i,j} = \alpha \mathrm{E}\left\{x_i x_j\right\}$$

- Total orbital-wise energy can then be written in matrix form:

$$J' = \frac{1}{2} x^T \mathrm{E}\left\{x x^T\right\} x$$

- Correspondingly for positive charges u_j (nuclei); forces are now attractive rather than repulsive

$$J'' = -x^T \mathrm{E}\left\{x u^T\right\} u$$

- For total energy one has

$$J(x,u) = J' + J'' = \frac{1}{2} x^T \mathrm{E}\left\{ xx^T \right\} x - x^T \mathrm{E}\left\{ xu^T \right\} u$$

  Here it is assumed that effects of the nuclei are quantized, and their effects are characterized by photon distributions determined by the relative locations of the atom nuclei
- The above J is exactly the same cost criterion that was derived for ordinary (neo)cybernetic systems!
- Resulting assumption: Thus, the charge distribution along the molecule (molecular orbital) is given by the principal components of the correlation matrix $\mathrm{E}\left\{ uu^T \right\}$ of photons carrying the nucleic interactions

Comparison to traditional theory

- Normally one has an (unsolvable) infinite-dimensional problem of eigenfunctions (time-independent formulation)

$$-\frac{h^2}{8\pi^2 m}\frac{d^2}{dx^2}\psi(x)+V(x)\psi(x)=E\psi(x)$$

Linear PDE – problem only with boundary conditions!

- Now there is only the finite set of nuclei being studied – one has a finite-dimensional eigenvalue/eigenvector problem

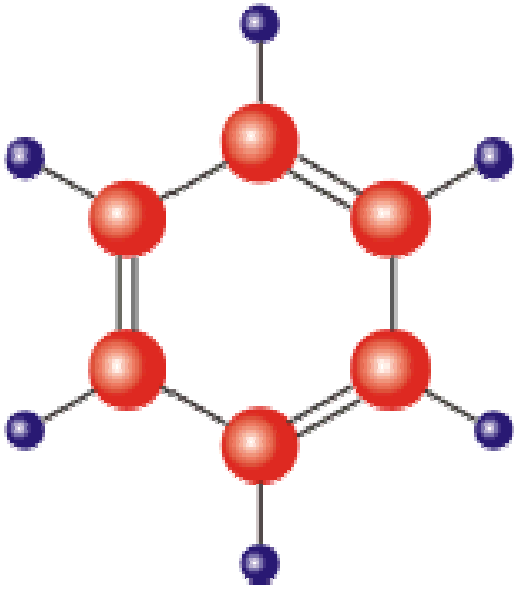
$$(V-V_0)\psi_i=\lambda_i\psi_i$$

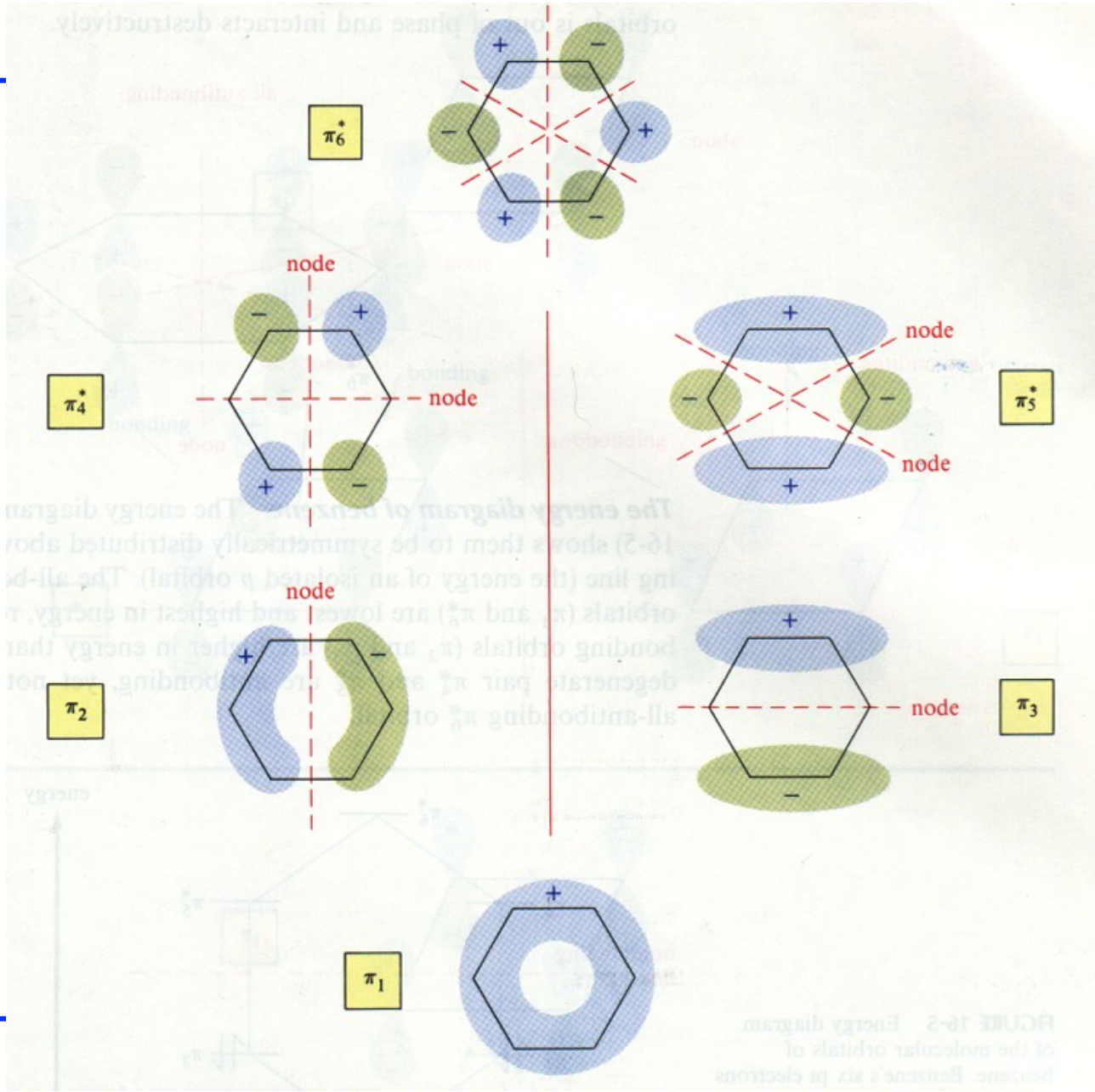
- Assumption: Because of the nature of electrons, they cannot be located in various energy levels simultaneously – eigenvalues become distinguished

- The above result is closely related to the *Hückel method*, where the molecular orbitals are (approximately) determined in a rather qualitative, graph theoretic way
- Molecular orbitals are interesting because the chemical properties are determined by the charge distribution = how the molecule is ”seen” by the outside world
- For example – if $\mathrm{E}\{uu^T\} = \begin{pmatrix} 1 & 0.5 & 0.1 \\ 0.5 & 1 & 0.5 \\ 0.1 & 0.5 & 1 \end{pmatrix}$ *chain of three?*

then $\psi = \left(\begin{array}{c|c|c} 0.52 & -0.71 & 0.48 \\ 0.68 & 0 & -0.73 \\ 0.52 & 0.71 & 0.48 \end{array}\right)$ and $\Lambda = \begin{pmatrix} 1.76 & & \\ & 0.90 & \\ & & 0.34 \end{pmatrix}$

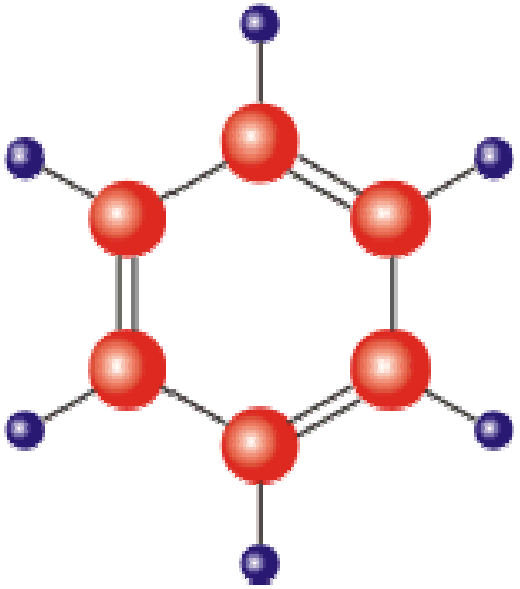
- Traditional view of orbitals in the benzene case

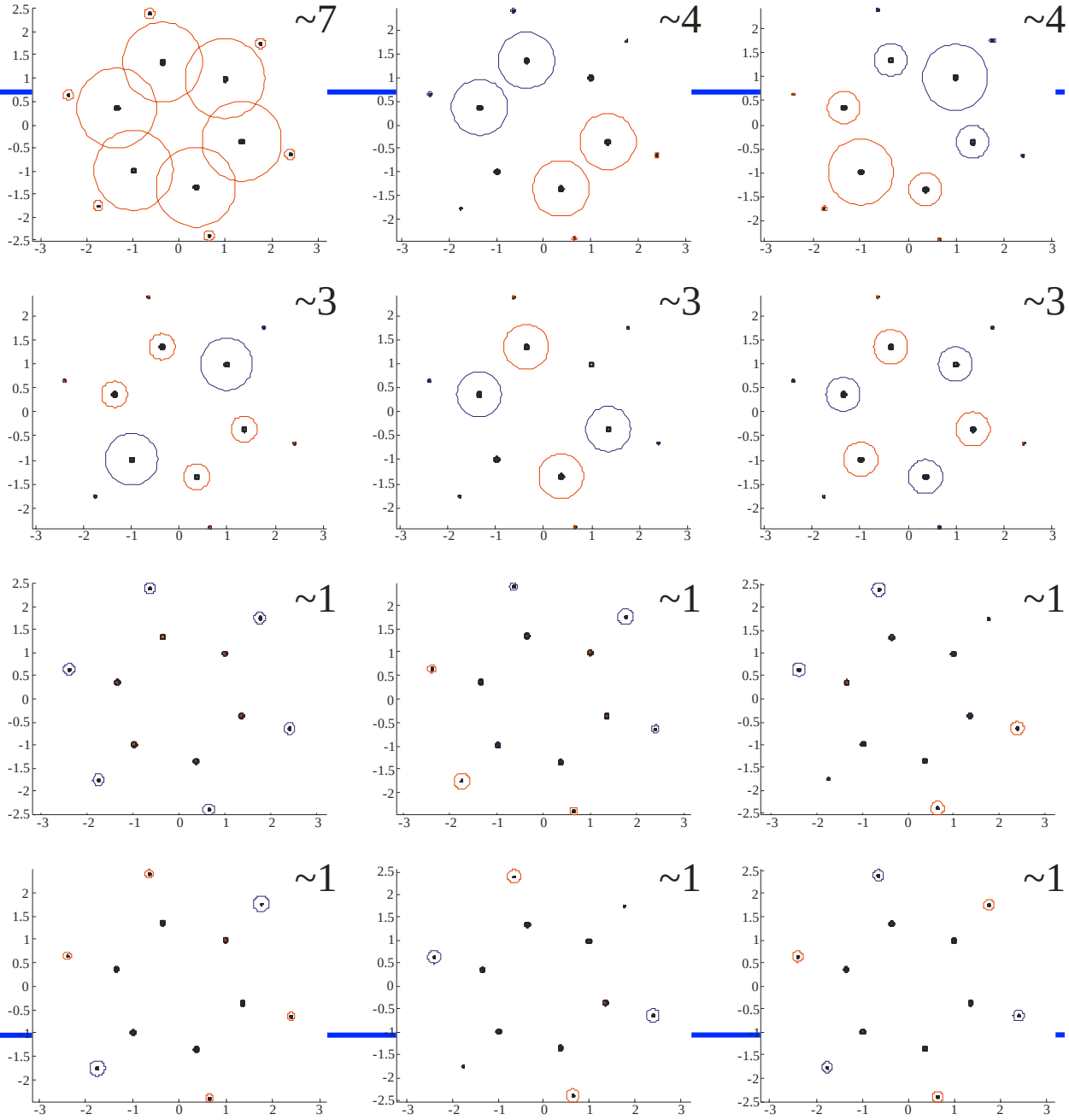




HELSINKI UNIVERSITY OF TECHNOLOGY
Control Engineering Laboratory
Cybernetics Group

- ”Cybernetic orbitals” for benzene





HELSINKI UNIVERSITY OF TECHNOLOGY
Control Engineering Laboratory
Cybernetics Group

- However, the complete solution of the Schrödinger equation is time-dependent:

$$\psi(x,t) = \psi(x)\mathrm{e}^{\mathrm{i}\ 2\pi Et/h}$$

- In our discretized case, one has

$$\psi_i(t) = \psi_i \sin\frac{2\pi\lambda_i t}{h}$$

- The energy eigenvalue λ_i determines the oscillation frequency of the orbital
- Emergent affinity = integral over time: Different orbitals do not interact
- Possibility of characterizing of atoms within a molecule!

- If one defines ”fingerprints” of atoms as

$$\Psi = \begin{pmatrix} \Psi_1 & \cdots & \Psi_n \end{pmatrix} = \begin{pmatrix} \psi_1^T \\ \vdots \\ \psi_n^T \end{pmatrix}$$

  one can write their mutual affinity as

$$\Psi_i^T \Lambda \Psi_j$$

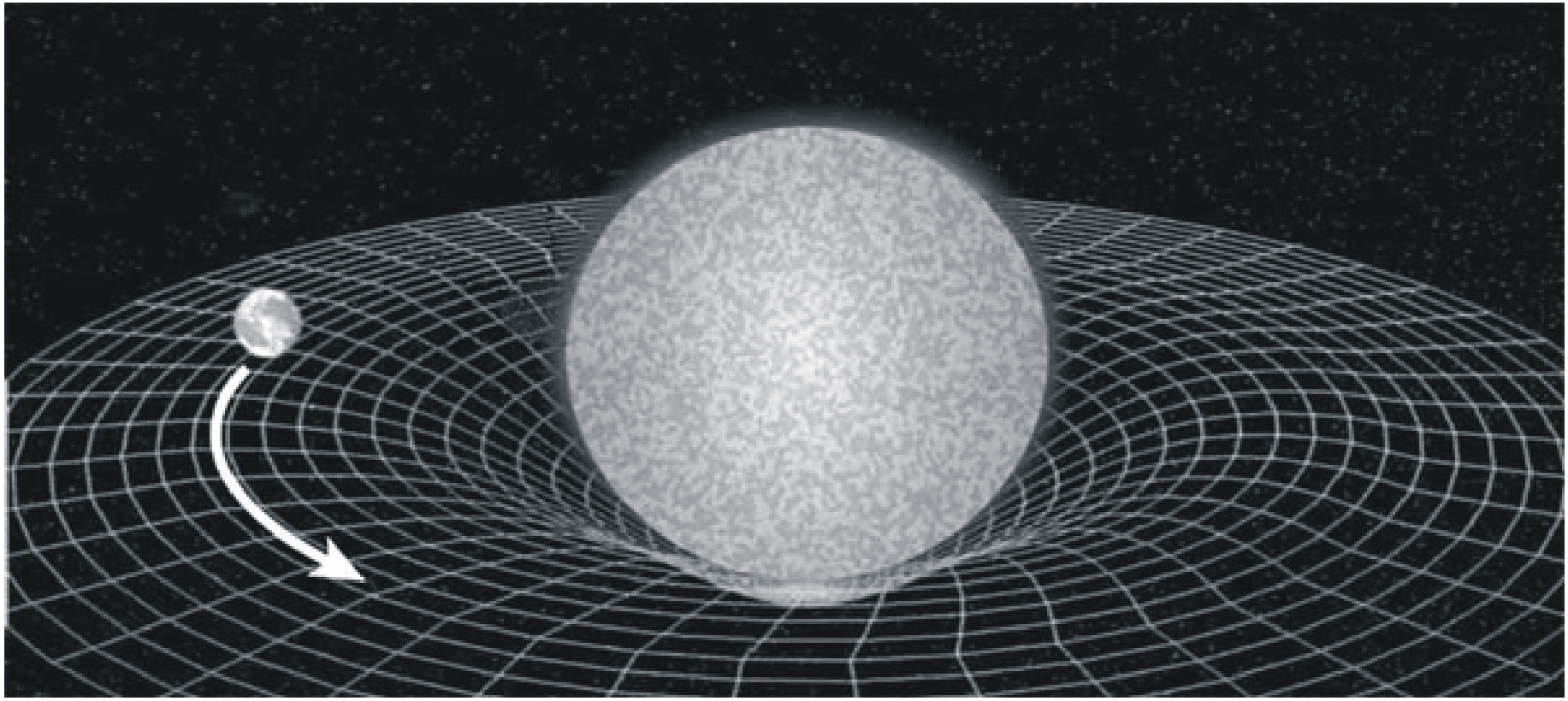
- This gives a unifying view over van der Waals bonds / hydrogen bonds + covalent bonds?
- Understanding of affinity between atoms i and j = contribution to protein folding, and activation energies?
- Infinite number of possible energy levels – infinite number of different affinity structures

Compare to the questions in the beginning:

- When affinities among atoms in a molecule are known, one can understand why different parts of the molecules become attached – explanation to protein folding?
- A related mysterious process is RNA splicing: The same DNA is expressed in different kinds of messenger-RNA because of splicing – the same explanation?
- If separate molecules synchronize the vibrations in their orbitals, their attraction patterns can also become infinitely complicated – explaining the diversity of protein functions?
- Further, as an enzyme molecule is attached to another molecule, the whole orbital structure is changed – thus altering the activation energies in other parts of the molecule

”Life, Universe, and Everything”

- ... How about the Universe and the string theories?
- Is Universe also an adapting elastic system?



Conclusion: About cybernetic systems

- Cybernetic system is a complex system that is characterized by dynamic equilibrium among opposing tensions
- The balances characterize dynamic attractors that are visible in the data and thus relevent in that domain
- Interacting systems are reactive, controlling each other, the overall dependencies becoming pancausal
- During evolution (natural or not) the controls become more and more stringent and the overall system becomes stiffer
- Final result: Degrees of freedom are eliminated

- The same principles apply to many different kinds of systems even if the phenomena cannot be explicitly quantified

Overlapping/interacting systems at TKK

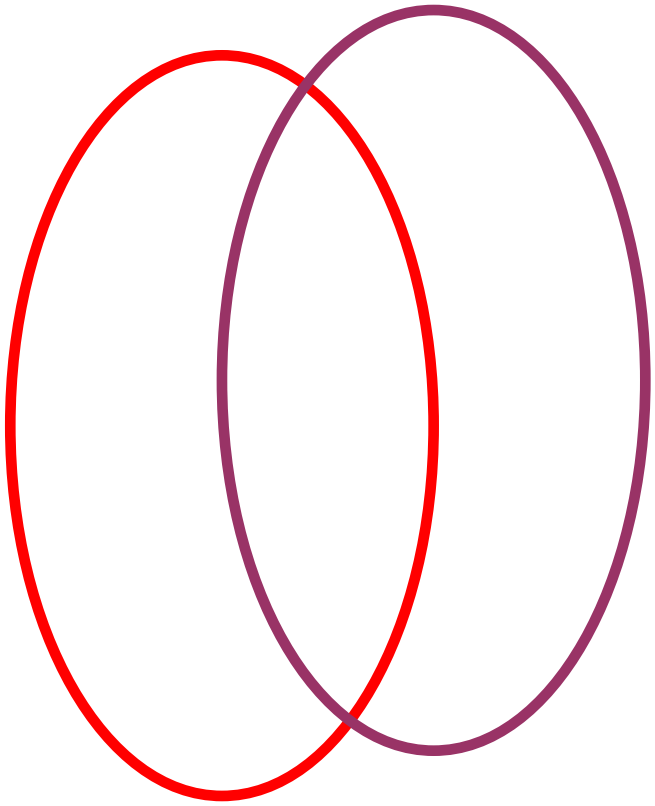
- There are coexisting consistent cybernetic subsystems
- Internal tensions keep systems ”alive” in dynamic balance

Scientific view

Criteria: Anarchy, non-formality, non-optimality

Emergent structures: Invisible ”hierarchy in substance”

Functioning: making questions, answering them



Administrative view

Criteria: Efficiency, optimality, money, formal frameworks

Prior structures: Organizational hierarchies

Functioning: Courses, projects, meetings, strategies

Cybernetization is not necessarily a good thing!

- Trends in working life:

1. Towards better understanding of the system and gaining more information (input)
   - Supervision of working time, questionnaires, more paper work
   - Terminology: ”Transparency”, ”efficiency”
2. Towards more efficient exploitation of information (control)
   - Expansion of administration, new ”planners”, organizational changes
   - Terminology: ”Near-boss”, ”developmental discussions”, ”competitiveness”, ”strategies”, ”missions and visions”

- Result: Freedoms/diversity explicitly eliminated
  - Is this not the cybernetic destiny? Is there any alternative?
  - In a research institution, *there should be*

- A scientific system is a cybernetic system, consisting of a population of independent actors = researchers
- Thus, scientific system is a control system, evolving towards better elimination of variability
- As the scientific system becomes ”better controlled”, there are stronger tensions in terms of competition
- A *paradigm* determines ”correct” ways to do research, defining standard problems and methods – *standard science*
- *Measurement: Evaluations, impact factors, peer reviews*
- *Control in terms of funding*
- One has to actively struggle against cybernetization!?