Belousov-Zhabotinsky Reaction and Pattern Formation in the Distributed Systems

Lecture 22
In this lecture we will continue the study of spatio-temporal patterns emerging in distributed systems. As a reminder from the previous lecture, the first necessary condition of the emergence of spatial structures is the fact the size of the system $L$ should greatly exceed the characteristic diffusion length pertaining to this system $L_{\text{diff}}$. Satisfaction of this condition requires us to convert models from easily analyzable ODEs to hard to work with PDEs – reaction-diffusion equations. But for that price the buy the chance to observe in the models the most fascinating and awesome patterns.

Perhaps it may come as a surprise that although our final goal is to study biological systems, we start with pattern formation in the chemical systems. There are two valid reasons why we proceed in this order. One of them is scientific and the other is methodological. Indeed, speaking technically, biological systems are chemical systems of outmost complexity and the formation of patterns in biological systems in many cases can be ascribed to chemical mechanisms. Later in the course we will specifically consider the situations in which chemical formalism is insufficient to explain the formation of biological pattern. What perhaps is even more important, is that chemical systems have been more extensively studied due to their easier experimental access and degree of control. Chemical systems are also useful to demonstrate the development of the mathematical apparatus which is routinely used in all areas of pattern formation.

So we continue our ascend of the complexity mountain working our way up from more intuitive to less intuitive systems, phenomena and mathematical formalism.
Belousov-Zhabotinsky reaction – the history of discovery

• 1951 – Belousov discovers chemical reaction showing temporary oscillations of color – paper is rejected as non-scientific
• 1964 – the work is continued by A. Zhabotinsky who presents the reaction to the Western scientific community
• 1972 – Field, Körös and Noyes present first model of the reaction mechanism (FKN)
• 1980s – 1990s the BZ reaction is widely studied in CSTR and spatially distributed conditions. Waves, Turing patterns and various over nonlinear phenomena are discovered.

Belousov-Zhabotinsky reaction rightfully occupies the central place in the area of nonlinear chemical phenomena and pattern formation. There is hardly any other reaction of no practical use in the industry or medicine that has been so thoroughly studied for nearly 50 years. Everything started from the observation of a Soviet chemist Boris Belousov that certain mixture of chemicals containing citric acid and bromate shows periodic changes in concentrations of species. His paper was rejected on the premises that chemical reactions cannot oscillate, otherwise that would violate the laws of thermodynamics. Only in 1959 he managed to publish a brief note in the proceedings of a totally irrelevant medical conference. Fortunately, other Soviet physical chemist, Anatoly Zhabotinsky, noticed the work and modified the reaction by changing some reactants and made the appearance of the phenomena more dramatic by achieving bright, easy to observe colors. So what did these scientists observe? Let us have a look at a short mpeg movie...

Belousov died in 1970 and never enjoyed the triumph of his discovery. At that same time, the results of the work were presented by Zhabotinsky at the conference in the US and the new era of scientific exploration began. One of the most prominent contributors to the area is Richard Field who together with Koros and Noyes proposed the first mechanism for the BZ reaction and ever since that has been studying the reaction proposing two more elaborate and detailed mechanism for the reaction.
So what is actually BZ reaction. Here is one of the latest views on the detailed mechanics of the reaction from Dick Field’s 2001 paper (Faraday Discussions, v. 120, pp. 21-38). Some 48 individual reactions fall into several categories and involve complex interplay of many inorganic and organic molecules of which malonic and bromous acids. The reaction is usually catalyzed by the ions of cerium. The oscillations in this extremely complex system arise due to the interplay of autocatalytic positive feedback and time delayed negative feedback. Needless to say that you cannot see that from the full mechanism represented by this table.

Before plunging into any mathematics in order to model this system, let us ask ourselves the question that was raised by Belousov’s reviewers: Can chemical systems oscillate? Indeed, oscillation assumes the return of the system in the same state with some period of time. How does that work together with the dissipative nature of chemical reactions. If we try to run the BZ reaction in the isolated system like that beaker in the previous slide, we will find that after a few periods the oscillations will subside. Why? This is because the system will run out of supplies and get intoxicated by the products. In other words only a subsystem of the entire system shows periodic oscillations provided that there is a strong dissipation in the supporting part of the system. To maintain the oscillations we would have to constantly pump the system with new “fuel” and remove the products of some "non-oscillatory" reactions.
FKN mechanism and Oregonator

\[ X = HBrO_2, Y = Br^-_\text{aq}, Z = Ce^{4+} \]
\[ A = BrO_3^-_\text{aq}, P = HBr \]

The simplified core of the FKN mechanism can be represented by five reactions where only concentrations of \( X,Y,Z \) are variables while \( A \) and \( P \) are parameters (fixed by the experimental conditions).

The system of kinetic equations for this system introduced by Field et al in 1972 is known as **Oregonator**:

\[
\begin{align*}
    A + 2X & \rightarrow 2P \\
    X + Y & \rightarrow X + P \\
    A + X & \rightarrow 2X + 2Z \\
    2X & \rightarrow A + P \\
    Z & \rightarrow 0.5Y
\end{align*}
\]

\[ k_1 = 5 \cdot 10^{-5} \]
\[ k_2 = 2 \cdot 10^{-4} \]
\[ k_3 = 8 \cdot 10^{-4} \]

\[
\begin{align*}
    \dot{x} &= k_1ay - k_2xy + k_3ax - k_4x^2 \\
    \dot{y} &= -k_1ay - k_2xy + k_5z \\
    \dot{z} &= 2k_3ax - k_5z
\end{align*}
\]

After one of possible non-dimensionalization of the system it becomes:

\[
\begin{align*}
    \varepsilon \dot{x} &= qy - xy + x(1 - x) \\
    \delta \dot{y} &= -qy - xy + z \\
    \dot{z} &= x - z
\end{align*}
\]

Let us now have a closer look at the mechanism proposed by Field and co-workers in 1972. It still contains more than 20 reactions but the “core” of the system can be represented by only five species shown on the slide. If you look at the kinetic scheme of the reaction you will find many strange things like transformation of one type of species, e.g., \( Z \) into \( Y \) with some arcane non mass action law stoichiometric coefficient \( \frac{1}{2} \). This is because many reactions are lumped here to create this rather artificially looking system. As we discussed on the previous slide, to ensure the “right” dynamics of the subsystem \( (x,y,z) \) of interest, we must assume that concentrations of other species are controlled by the flows of reagents. If the now explicitly write down the system of ODEs describing the dynamics of these three variables, we will get the famous Oregonator system proposed by Field’s group. There are several flavors of this system depending of the way used to non-dimensionalize the system. Following Murray, we will use it the form shown on the slide.
Reduction of Oregonator to 2D system with relaxation

The existence of small parameters at the time derivatives allows for the reduction of system dimensionality. Let us assume that $x$ is in a quasi-stationary state and thus $dx/dt = 0$

$$0 = q y - xy + x(1 - x)$$

$$\delta \dot{y} = -q y - xy + z$$

$$\dot{z} = x - z$$

After first reduction we obtain 2D system with small parameter $d$. Analysis of this system demonstrates the phenomenon of relaxation oscillations.

Further reduction of the system will result in a single highly non-linear equation for $z(y)$

$$y = y(z) \Rightarrow x = x(z)$$

If now look carefully at the resulting system we will see the presence of two small parameters epsilon and delta multiplying time derivatives for $x$ and $y$. Presence of such parameters is indicative of phenomenon of time scale separation and allows us to reduce the system of equations explicitly choosing which time scale we are planning to consider. As epsilon is the smallest parameter, we will start with variable $x$. Indeed, smallness of epsilon indicates that the time dynamics of $x$ is faster than that of $y$ and considerably faster than that of $z$. Thus the long term dynamics of $x$ will be confound to a so-called slow manifold provided that it is stable. We will shortly see when this manifold does and does not satisfy this condition. Now, assuming $dx/dt = 0$ gives the equation defining this slow manifold for $x$ which we can express as a function of $y$, $x = x(y)$. After that we are down to two equations for $y$ and $z$ only again with a small parameter multiplying the time derivative of $y$. We could continue to reduce the system by assuming $dy/dt = 0$. This will bring us to a single equation for $z$ with highly nonlinear right hand side. This is a general situation arising from the application of method of separation of time scales which is in chemical kinetics is known as the quasi-stationary approximation. By using it, we reduce the dimension of the system while paying for it by increasing the non-linearity of the system.

We however will stop on the system of two equations and carefully explore its dynamics. The shape of the nullclines is as shown on the figure. What does the presence of small parameter delta mean? It means that trajectory started anywhere on the $(y,z)$ plain will quickly reach the stable part of the slow manifold – this is exactly the red nullcline – and then will slowly follow it. It can be shown that only those parts of the nullcline which have positive derivative $dz/dy$ are stable, while the negatively sloped part is absolutely unstable. Therefore the intersection of the nullclines on the unstable branch of the slow manifold results in the unstable stationary point (which is fact an unstable focus). Putting this all together we will come to the conclusion that the system must have a limit cycle with two slow and two fast phases corresponding to the “slip” of the trajectory from the stable slow manifold. The phenomenon of oscillation with such uneven phases characterized by two very different time scales is called relaxation oscillation.
Consider now again generic 2D system with cubic nonlinearity in the $dx/dt$ nullcline but with small parameter at $dx/dt$, or generalized FitzHugh-Nagumo model:

$$\epsilon \dot{x} = f(x) - y$$
$$\dot{y} = y - ax - b$$

Of particular interest is the combination of parameters resulting in a single stable state close to the “knee” of the nonlinear nullcline:

Of particular interest is the combination of parameters resulting in a single stable state close to the “knee” of the nonlinear nullcline:

Now let us step aside from the BZ system and its simplified representation by the Oregonator model and have a broader look at the two-variable systems with cubic nonlinearity. Perhaps you remember that we once visited the layout of nullclines shown on the figure. If we reduce the angle of the linear nullcline, we will have three intersection points – two stable ones and one saddle and the system will be bistable. The only difference with that consideration is that we now in addition assume that the variables $(x,y)$ have different time scales which we explicitly introduce with the epsilon. On the previous slide we explored what happens when we have a single intersection point of the nullclines on the unstable part of the nonlinear nullcline. Now let us what happens if we move the linear nullcline past the “knee” of the S-shaped nullcline into the stable area without forming more intersection points as shown in the figure. Formally we’ll get a single stable state and nothing interesting. However, if the force the system to leave the stable state by applying perturbation of finite size as shown by the red arrow in the figure, we will get a very non-trivial response. Instead of plain relaxation to $S$, the system will execute a large loop shown in blue, very much like one period of the relaxation oscillation discussed before. The dependence of variables $x$ and $y$ on time is shown on the right.
Waves in excitable media

Provided some additional conditions are satisfied, the spatially distributed system with excitable dynamics will support waves:

Let us now assume variable y does not diffuse and we substitute time variable by “fast” time of the variable x:

With this transformation the system becomes:

\[ \dot{x} = f(x) - y + D_x \Delta x \]
\[ \dot{y} = \varepsilon (by - ax) \]

Now we repeat the trick of transforming variables \((x,y)\) to a single wave variable \(z=x-ct\) to find a potential wave solution:

\[ 0 = f(x) - y + D_{xx}'' + c x' \]
\[ cy' = \varepsilon (by - ax) \]

This system of two differential equations just like in the case of bistable systems can be transformed into a system of first order equations:

\[ x' = w \]
\[ w' = -\frac{1}{D} (f(x) - y + cw) \]
\[ y' = -\frac{\varepsilon}{c} (by - ax) \]

What if we now spatially distribute this system? It is always a good idea to start with one spatial dimension. Intuitively, it is obvious that if we somehow managed to create excitable response in one of our imaginary cubic-shaped reactors connected by diffusive coupling, this element will serve as ignition for the neighboring elements. By the time the dynamics in this cube returns to S the two neighboring elements will be already on their excitable loops. In turn, these elements will serve as ignition to their neighbors and so on. Finally we have a propagating pulse consisting of two fronts — ignition and extinction — following each other in space and time.

The mathematical description for excitable pulses is considerably more involved than that for the bistable systems and we will not go into it in depth. It is however worth while to explore some the analogies with the mathematical formalism developed in the previous lecture. The main trick here is the same — searching for the wave solution in the form of functions depending on the wave variable \(z=x-ct\). To make things simpler, we will assume that only one variable, the fast one, diffuses. This in fact corresponds to the FitzHugh-Nagumo model for the neural pulse propagation. In this case we only have one equation with the second derivative over \(z\) and the full system can be converted into the system of ordinary differential equations in 3D space. Let us explore it in more detail.
Waves in excitable media

Just like in bistable case, transition to variables \((x,y,w)\) changes the stability of the stationary point \((0,0,0)\) from stable node to a saddle:

Depending on the initial conditions, the wave will either develop into a propagating one or into a dying out pulse.

The initial condition is not the only requirement for propagation. It can be shown that speed of both stable and unstable solutions depend on \(e\).

The waves in the original spatially distributed system will correspond to homoclinic trajectories in this system. In fact, there are two: stable and unstable (red).

Continuing the analogy with the bistable system, out of all possible trajectories of this \((x,y,w)\) system we are looking for a very special one which originates at the stationary point \(S'\) (this is the image in \((x,y,w)\) of the state \(S\) in the \((x,y)\) ODE system) and eventually arrives to \(S'\). This is a very special type of trajectories which only exist under special conditions and are called homoclinic orbits. Technically, they are limit cycles of infinite period. Remember now that the “time” in the \((x,y,w)\) system is in fact a wave variable \(z = x-ct\). Therefore, this exotic trajectory in the \((x,y,w)\) space corresponds to a propagating excitation pulse in the corresponding RDE.

Without going deeply into the matter, I will only summarize the main conclusions. It happens so that the system of three equations on the previous slide has exactly two different homoclinic orbits for fixed value of the parameter epsilon. One corresponds to the actual pulse which can propagate without change in profile and the other can not be observed as it is absolutely unstable. More correctly, the two solutions exist only for a range of epsilon less then the critical value. At this value, both solutions come together in a saddle-node bifurcation for closed trajectories and disappear. The important conclusion from this mathematical fact is that no excitable waves can propagate in a system with insufficient difference in time scales.
BZ medium can be simultaneously oscillatory in well-stirred reactor and excitable in a thin 2D layer open to oxygen.

Excitable waves in the BZ system are mostly observed in the conjunction of so-called target patterns.

To induce these patterns, one needs to initiate oscillations at a point in the medium – excitable dynamics is not self-sustainable unless initiated in the periodic spatial domain.
Spiral waves – phenomenological description

If excitable wave front is mechanically broken, the free end of the front will develop into a so-called *spiral wave* as shown in the figure.

Spiral waves have a number of characteristic properties:

• Their wavelength (or *pitch*) is considerably shorter than that of a target.

• Spiral waves tend to suppress other spatial patterns like targets in a homogeneous system – this has a very important consequence in cardiac arrhythmias (see later).
For nearly 30 years spiral waves have been studied in the Belousov-Zhabotinsky system under all kinds of conditions and external perturbations. Changing the concentrations of chemicals, parameters of reactor feeding and perturbations by light, it is possible to observe a staggering variety of phenomena which include such exotic manifestations as meandering movement of the spiral wave cores and coexistence of spiral waves with labyrinthine patterns as shown in the figure. The latest work I was able to find on spirals in BZ published by Vanag and Epstein in PNAS (2003, V. 100, p. 14635) involves segmented waves in a water-in-oil microemulsion. I confess to spending four years in spirology myself while doing my doctoral work on spiral waves in media with complex periodic and chaotic dynamics.

By now you are perhaps firmly convinced that chemical pattern formation is extremely exotic area far away from any biological applications. This is not exactly the case as we will see on the next slide.
Where can we find spiral waves in the live nature? Surprisingly, in a whole lot of system of very variable size. Here you can see spiral waves of electric activity on the chicken retina, cAMP waves formed by moving colonies of Dictiostellium discoideum and my personal favorite, a plant with a spirally growing leaves in the Singapore Botanical garden. One of the most important applications of spiral waves in biology has been so far pathologic formation of spiral waves of activity in the heart. In the norm, our hearts rely on a stationary pacemaker to create a circular wave of excitable nature which propagates through the tissue causing it to constrict. Under pathological conditions, for example when due to the ischemic damage some parts of the tissue are no longer excitable, this circular wave brakes with inevitable formation of a couple of spiral waves. The bad thing about them is that unlike the circular wave, spirals are self perpetuating patterns and once established will stay in the heart. This results in the formation of potentially lethal atrial fibrillation. Despite a common perception that spiral waves require large domains for their existence, spiral waves can exist on really microscopic scale. Recently it has been shown that spiral waves can exist on the surface or inside of individual cells as small as common neutrophils.
What to take home

• Pattern formation in chemical systems provides phenomenologically and methodologically rich foundation for modeling of complex biological phenomena.
• Spatio-temporal patterns observed in distributed chemical systems are mainly defined by the internal properties of these media.
• A variety of complex dynamical patterns can be classified into three main categories according to the media properties: switch-like (multistability), excitable and oscillatory.
• The most relevant for biological applications is the excitable dynamics. It is observed as propagating pulses of excitation in the form of plane waves, circular waves and target patterns and spiral (scroll) waves.
• Spiral waves are widely occurring patterns of dynamics in biological systems. They can be identified on all scales of biological organization from intracellular to ecological population.