Precise and Robust Regulation of Gene Expression Using Fully Automatic Optogenetic Feedback

Andreas Milias-Argeitis
(*joint work with Marc Rullan, Stephanie Aoki, Peter Buchmann and Mustafa Khammash*)

Dept. of Biosystems Science and Engineering
Laboratory of Control Theory and Systems Biology
ETH Zurich

Séminaire d'Automatique du Plateau de Saclay

12.11.2015

Outline

1. Preliminaries: molecular biology
2. Preliminaries: optogenetics
3. Motivation
4. Experimental system + setup
5. Controllers
6. Experimental results: tracking + disturbance rejection
7. Discussion

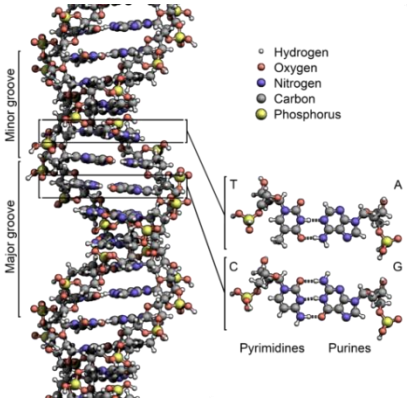
Outline

1. Preliminaries: molecular biology
2. Preliminaries: optogenetics
3. Motivation
4. Experimental system + setup
5. Controllers
6. Experimental results: tracking + disturbance rejection
7. Discussion

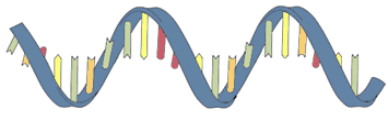
Molecular biology: absolute basics (1)

Molecular biology studies biological activity at the molecular level; more specifically the interactions between DNA, RNA and proteins inside the cell

DNA carries the genetic information for the development, functioning and reproduction of all living organisms. Information is organized in **genes**.

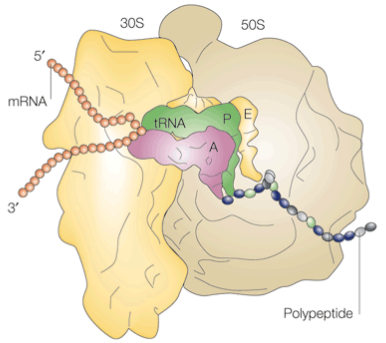


RNA is similar to DNA in composition, but carries out a large variety of biological roles. **mRNA** (messenger RNA) conveys genetic information from the DNA to the **ribosome**.



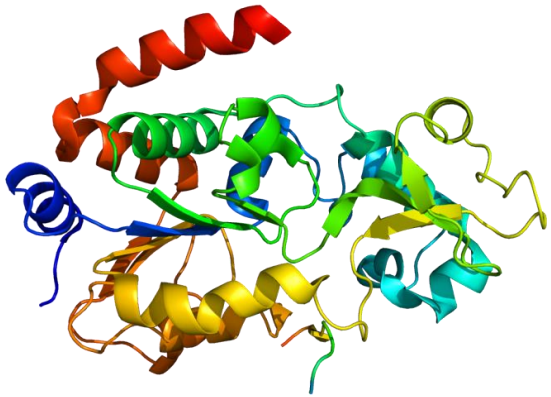
Molecular biology: absolute basics (2)

The **ribosomes** are the molecular devices that synthesize proteins inside the cell using the information coded in the mRNA



Nature Reviews | Molecular Cell Biology

Finally, **proteins** are the molecules that carry out almost every cellular function (metabolism, nutrient transport, signaling, DNA replication, RNA transcription, molecule transport etc.). They consist of long chains of **aminoacids**.

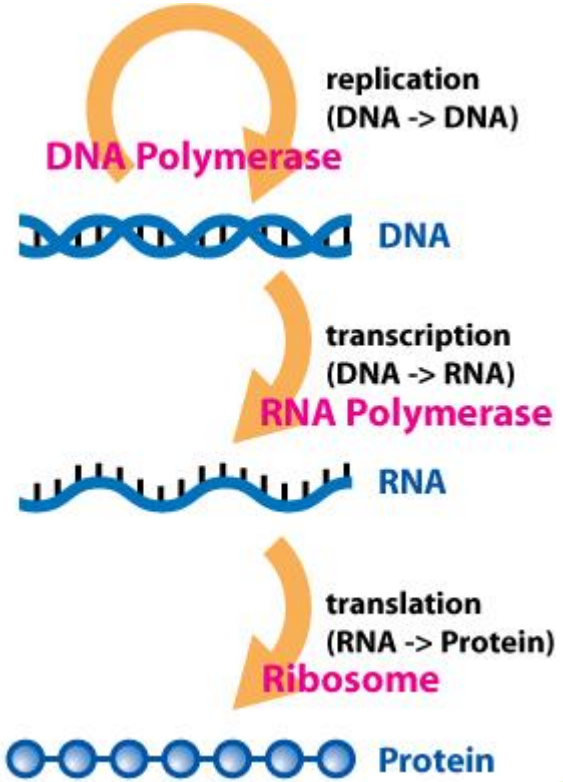


The central dogma

Or, how genetic information flows within a biological system:

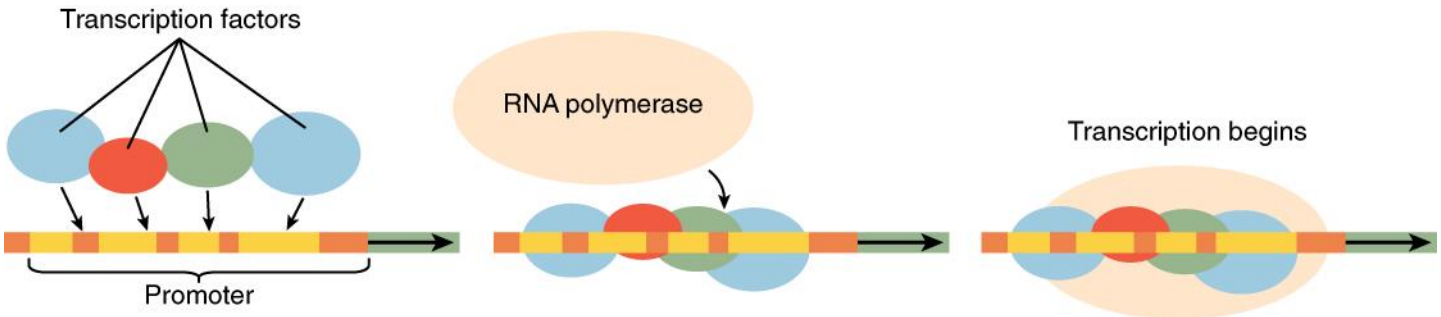
Information stored in DNA nucleotide sequence is *transcribed* into mRNA nucleotide sequence, which is in turn *translated* into protein aminoacid sequence.

The process of ***gene expression***



Transcription factors

- Not all genes are expressed inside the cell at the same time
- Gene expression is a very finely **regulated** process
- A set of DNA-binding proteins help initiate (or repress) transcription of a given gene: the **transcription factors (TFs)**
- Each gene has different TF **binding sites**. TFs cross-regulate each other's expression → **genetic regulatory networks**



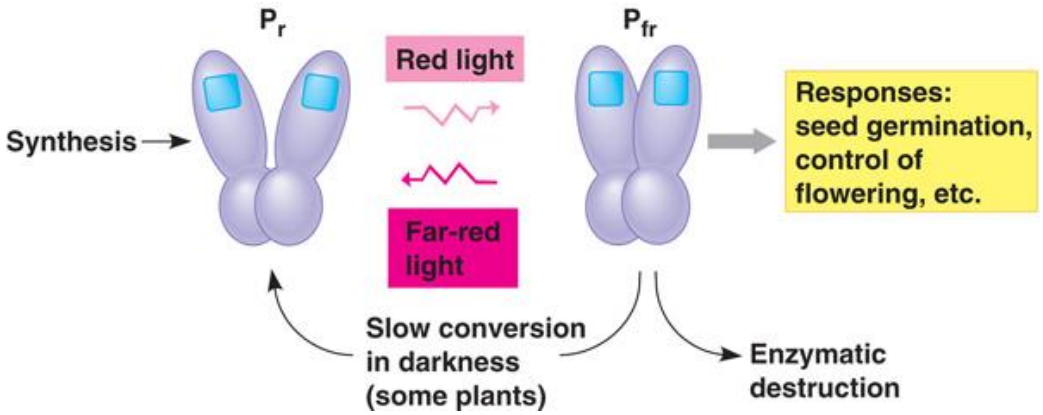
Outline

1. Preliminaries: molecular biology
2. Preliminaries: optogenetics
3. Motivation
4. Experimental system + setup
5. Controllers
6. Experimental results: tracking + disturbance rejection
7. Discussion

Optogenetics

A set of biological techniques involving the use of light to control cells in cultures, living tissues etc. that have been genetically modified to express **light-sensitive proteins**

Examples of light-sensitive proteins: photoreceptors in the eye; plant photoreceptors (phytochromes); bacterial photoreceptors



Optogenetics

Today there exist engineered photoreceptors that can carry out a large variety of protein functions, enabling the optogenetic control of:

- Neuronal activity
- Gene expression ← our work
- Cell signaling
- Protein localization
- Cell migration
- …

Advantages: rapid, targeted, low-cost and precise spatio-temporal modulation of protein function, low toxicity, no pleiotropic effects

Outline

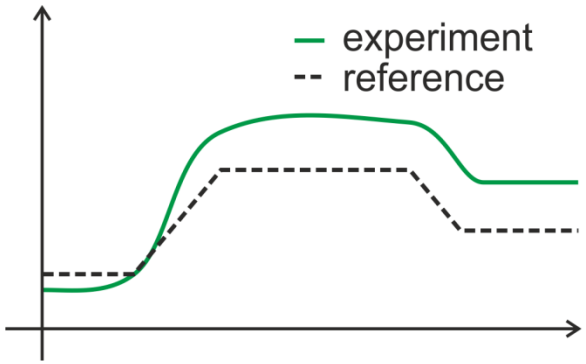
1. Preliminaries: molecular biology
2. Preliminaries: optogenetics
3. Motivation
4. Experimental system + setup
5. Controllers
6. Experimental results: tracking + disturbance rejection
7. Discussion

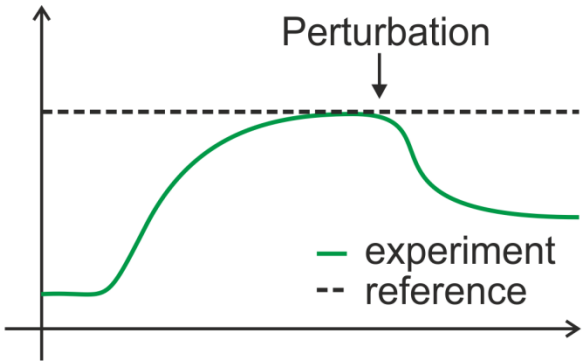
Motivation

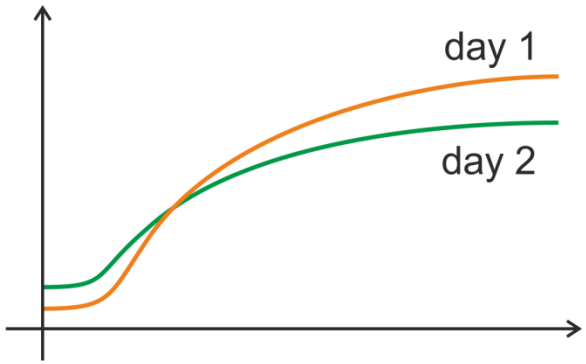
- Optogenetic manipulation of biological systems has a great potential, both for basic and applied biological research
- Optogenetic systems can be used either as “function generators” (e.g. for discovery) or “process controllers” (e.g. in metabolic regulation)
- In any case, they need to be able to operate robustly within complex and variable cellular contexts
- However, “optogenetic control” is not actually control – rather, it should be called a perturbation
- For precise and robust regulation, perturbations are not accurate enough: actual feedback control is necessary

Why is feedback necessary?

Example: control of gene expression







Model mismatch:

Long-term tracking becomes impossible with an inaccurate model

Perturbations:

Left uncompensated, can lead to substantial deviations from desired tracking target

Day-to-day variability:

Same input applied on different days results in different outputs

Outline

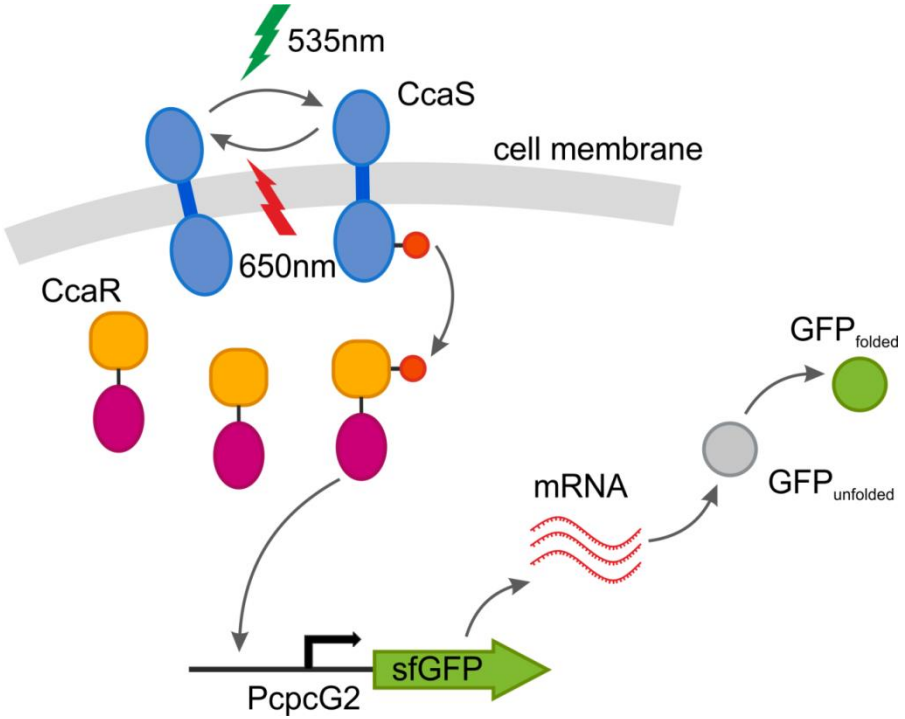
1. Preliminaries: molecular biology
2. Preliminaries: optogenetics
3. Motivation
4. Experimental system + setup
5. Controllers
6. Experimental results: tracking + disturbance rejection
7. Discussion

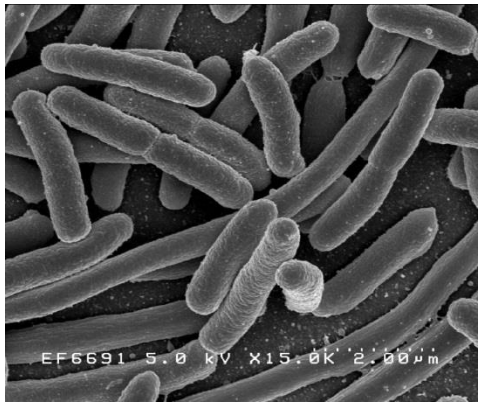
The optogenetic system

The CcaS-CcaR cyanobacterial two-component system

- Photosensor: CcaS
- Transcription factor: CcaR
- Output monitoring: Green Fluorescent Protein (GFP)

Implemented in *Escherichia coli*





ETH zürich

D-BSSE
Department of Biosystems Science and Engineering

Used equipment

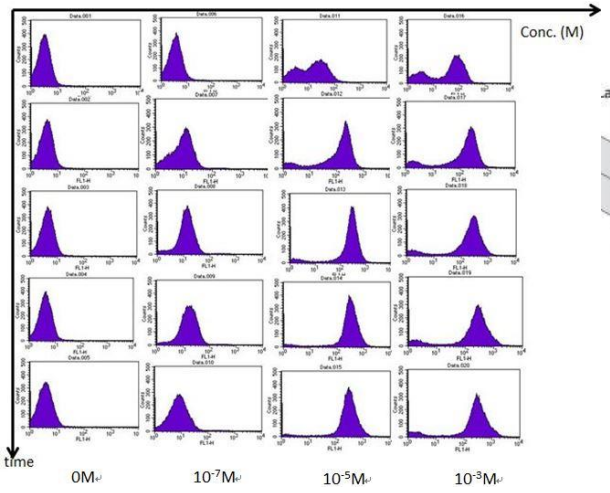
1. Turbidostat

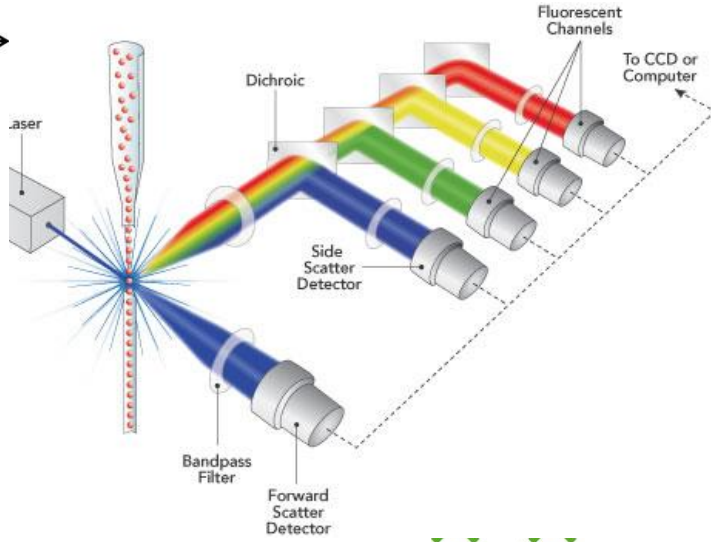
- Under most commonly used steady-state growth conditions, a bacterial cell divides every 20-60 mins.
- The number of bacteria in a liquid culture increases exponentially
- If left uncontrolled, the bacterial population will eventually deplete the nutrients in the culture medium and steady-state growth will stop
- The turbidostat maintains a constant culture tubidity by continuously measuring the culture absorbance and adjusting the inflow/outflow of medium

Used equipment

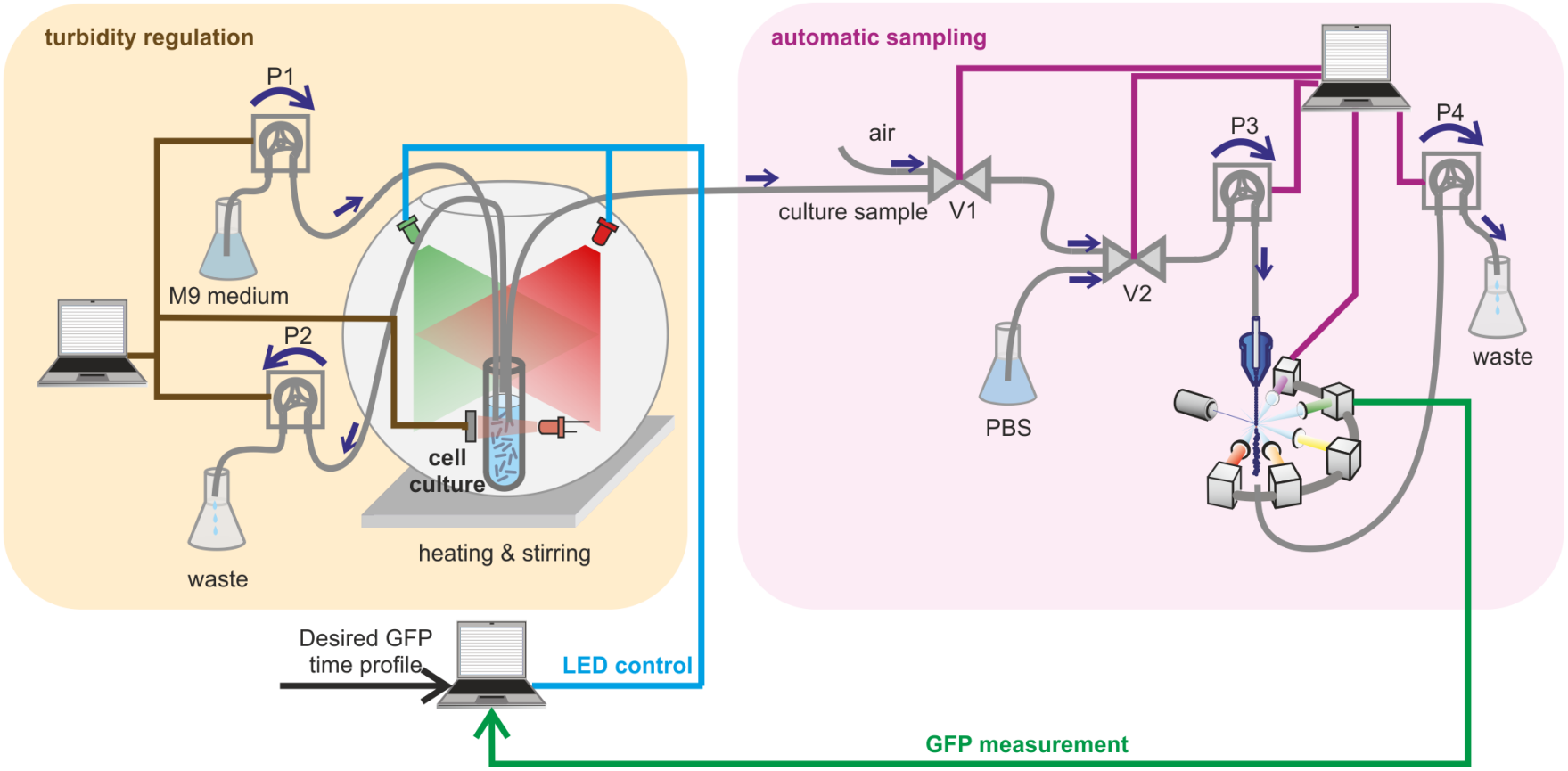
2. Flow cytometer

- Flow cytometry is a commonly used technology for analyzing cell properties (in our case, total cell fluorescence)
- Cells are labelled, suspended in a stream of liquid, excited by multiple laser sources and passed through a detection apparatus
- We thus obtain distributions of single-cell properties, such as fluorescence, volume etc.





Experimental setup



Outline

1. Preliminaries: molecular biology
2. Preliminaries: optogenetics
3. Motivation
4. Experimental system + setup
5. Controllers
6. Experimental results: tracking + disturbance rejection
7. Discussion

Feedback control of GFP expression

Objective: regulate the *mean* GFP concentration (fluorescence/volume) of an *E. coli* population in liquid culture by modulating the green light intensity for a fixed red light level

Controllers:

- Proportional-integral (PI)
- MPC with a fixed model + Kalman filter
- MPC with particle filter (state and parameter estimation)

Controllers tested on reference tracking + disturbance rejection

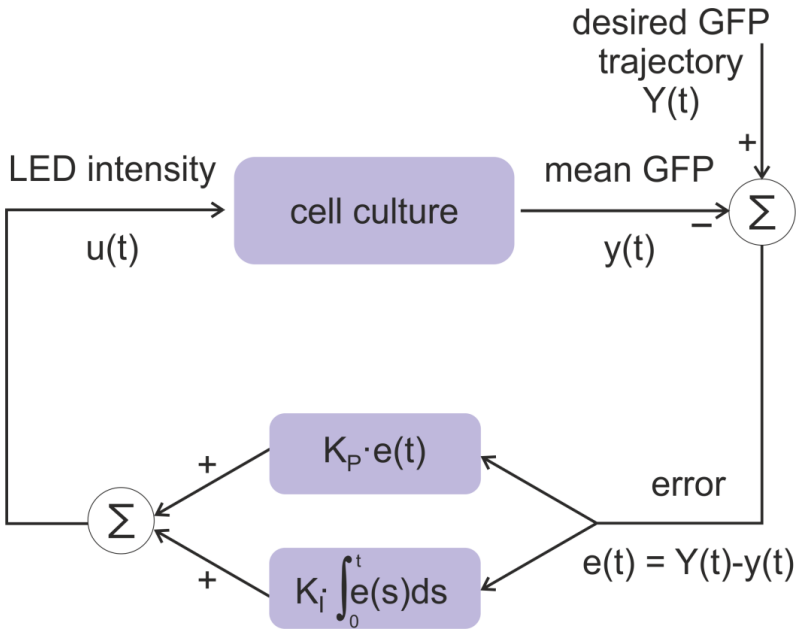
PI controller

Pros:

- Zero tracking error for constant references
- Rejection of constant disturbances

Cons:

- Careful tuning of two parameters (+gain-scheduling for perturbation rejection)
- Poor non-constant reference tracking



Discrete-time implementation (forward difference), sampling time T = 10 min.

PI controller

Further details:

- Initial guesses based on a black-box model fitted from characterization data. Then, trial and error.
- First tuned parameters for tracking a step reference (objectives: converge fast, avoid large overshoot)
- However, this parameter set was too “weak” for perturbation rejection (“static friction” effects)
- Definition of a second parameter set that is used after setpoint is achieved (to anticipate perturbations)
- Extra difficulty: system nonlinear, tuning changes with setpoint
- Simple, straightforward design

Basic MPC scheme

Basic scheme:

Given: a system model, tracking objective, input + state constraints, control horizon, cost function, initial state conditions

1. Determine input sequence that optimizes the cost over the control horizon using the system model
2. Apply the first sample of the input sequence
3. At next sampling instant: measure system output and estimate system state
4. Go to 1 (shift the control horizon one step ahead and repeat)

MPC features

- Convergence to the reference guaranteed when the controlled system is *linear* and the model *precisely known*
- Control of nonlinear systems is usually done without any guarantees on convergence (non-convex optimization → solver may fail!)
- More relevant for us: controller performance can deteriorate *a lot* when the used model is inaccurate
- Modeling inaccuracies are typically dealt with *robust MPC* (however, modeling assumptions on the inaccuracies are necessary)

Modeling details: original model

States: GFP mRNA (*r*), immature GFP (*p*), mature GFP (*g*)

Reactions:

Light-controlled

mRNA transcription + degradation $\varnothing \xrightarrow{b_{r0}+U(t)} r \xrightarrow{d_r} \varnothing$

immature GFP translation + dilution $\varnothing \xrightarrow{r} p \xrightarrow{d_p} \varnothing$

mature GFP production + dilution $p \xrightarrow{k_m} g \xrightarrow{d_p} \varnothing$

Equations:

$$\frac{d}{dt}\begin{bmatrix} r \\ p \\ g \end{bmatrix} = \begin{bmatrix} -d_r & 0 & 0 \\ b_p & -d_p - k_m & 0 \\ 0 & k_m & -d_p \end{bmatrix} \begin{bmatrix} r \\ p \\ g \end{bmatrix} + \begin{bmatrix} b_{r0} & b_r \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ U \end{bmatrix}$$

Fold-change model

We do not have access to absolute amounts of the species → further conversions are necessary

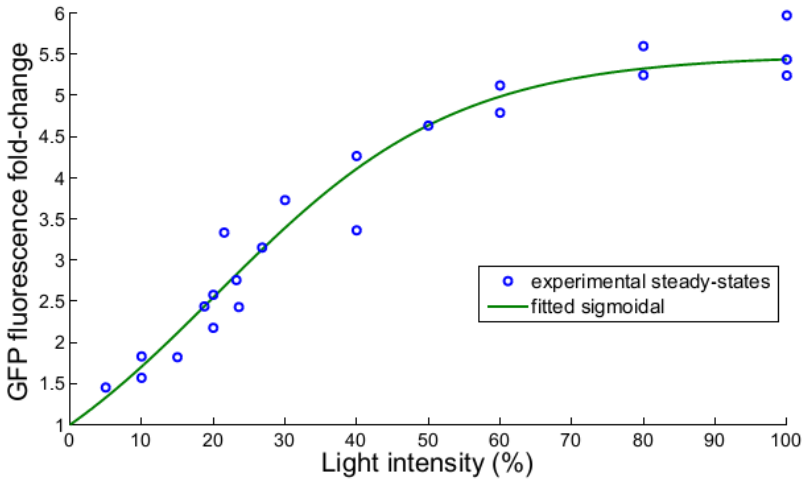
- Fold-change model (assuming initial equilibrium for a given input)

$$\frac{d}{dt}\begin{bmatrix} R \\ P \\ G \end{bmatrix} = \begin{bmatrix} -d_r & 0 & 0 \\ d_p + k_m & -d_p - k_m & 0 \\ 0 & d_p & -d_p \end{bmatrix} \begin{bmatrix} R \\ P \\ G \end{bmatrix} + \begin{bmatrix} d_r & b_r \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ u \end{bmatrix}$$

$$u = f(U)$$

↓

Connected through the **dose-response** curve



ETH zürich
D-BSSE
Department of Biosystems Science and Engineering

Final model

The fold-change continuous-time model was converted to discrete time (T_s = 10 min., piecewise constant input)

Further details:

- Addition of an uncontrolled input (modelling additive disturbances)
- One-step time delay to the controlled input (optogenetic system activity changes are not instantaneous)

$$X_{k+1} = A_d X_k + B_d \begin{bmatrix} 1 \\ u_{k-1} + d_k \end{bmatrix}$$

$$d_{k+1} = d_k$$

$$y_k = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} X_k$$

Simple MPC is insufficient

MPC is by design able to exploit the fact that the *short-term* behavior of the model matches that of the controlled system. Over *longer* time horizons, the mismatch is less significant.

The simple linear model cannot capture the nonlinear effects that follow the application of a light input, but does much better in the long run → the opposite of what is needed!

Additional problems:

- Day-to-day variability in cell behavior
- Non-additive perturbations

➢ Tracking performance severely degraded by large oscillations…

Adaptive MPC

Online state and parameter estimation is necessary:

- State estimation for MPC feedback
- Parameter estimation for counteracting parameter shifts due to day-to-day variability and disturbances (e.g. medium or temperature shift)

Solution: particle filtering:

- With the addition of the parameters to the estimated states, the model becomes *nonlinear*. Extended Kalman filtering becomes very difficult to troubleshoot
- Particle filters are known to converge to the true posterior distributions as the number of particles increases

Particle filter details

- Convert unknown parameters to states affect by zero-mean white noise
- For original system states: assume process noise dominates measurement noise (flow cytometry is quite accurate)
- At a given step: obtain posterior estimate of *current* (augmented) state given all measurements up to this step
- This is called the *Marginal Particle Filter* ([Klaas et. al])
- Necessary because parameter means must be able to shift over time. Classical Bootstrap Particle Filter failed at handling large parameter shifts.

The Marginal Particle Filter

Algorithm 1 The Marginal Particle Filter.

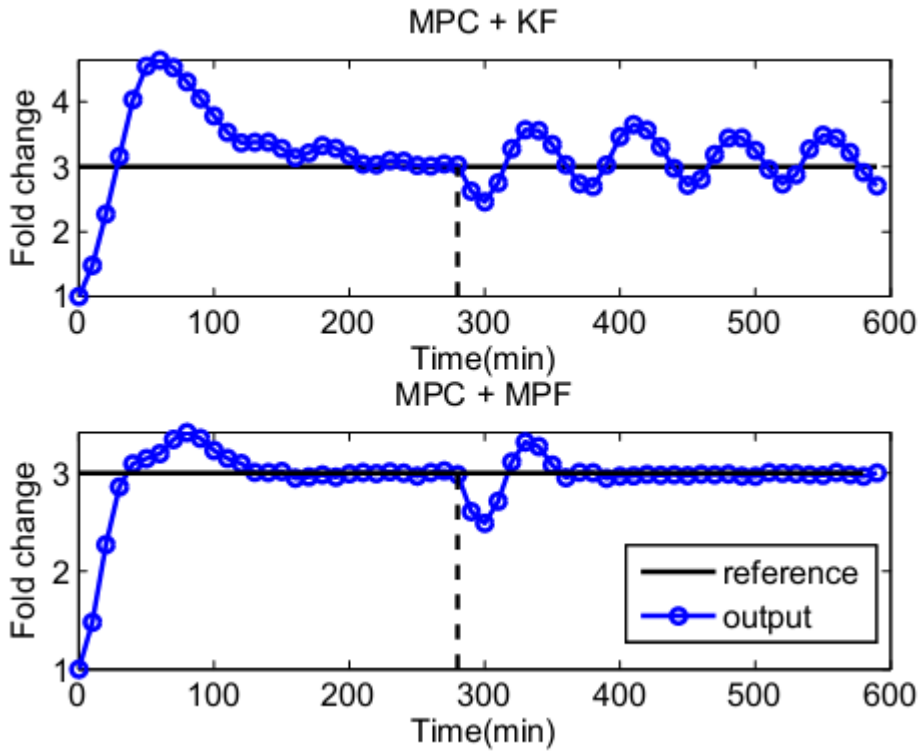
{*Setup particles*}
Draw P parameter particles from the prior: $\{\theta_0^p\}_{p=1}^P \sim \mathcal{N}(\theta^{nom}, \Sigma_\theta)$; generate P state particles $\{x_0^p\}_{p=1}^P$, all equal to $\begin{bmatrix}1 & 1 & 1\end{bmatrix}^T$.
for n from 1 to T **do**
  for p from 1 to P **do**
    {*Propagate state particles*}

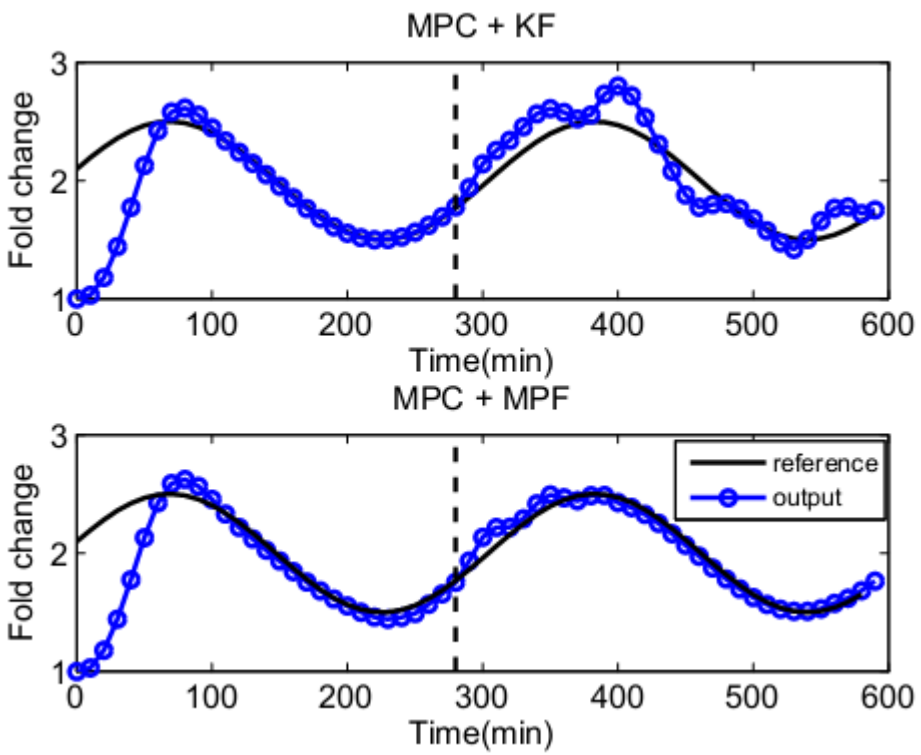
$$x_{p,n} = A(\theta_{p,n-1})x_{p,n-1} + \\ + B(\theta_{p,n-1})\begin{bmatrix}1\\ u_{n-1}+d_{n-1}\end{bmatrix} + w_n,$$
$$w_n \sim \mathcal{N}(0, \Sigma_s)$$
$$y_{p,n}^{pred} = Cx_{p,n}$$

  end for

{*Compute and normalize particle weights*}
$w_p = f(Y_{p,n}^{pred} | y_n^{meas}, \sigma_{meas}^2)$
$W_p = w_p / (\sum_{j=1}^{P} w_j)$
{*Resample states and parameters*}
Draw P state particles from $\{x_{p,n}\}$ and P parameter particles from $\{\theta_{p,n-1}\}$ according to weights $\{W_p\}$
for p from 1 to P **do**
  {*Perturb parameter particles*}
  $\theta_{p,n} = \theta_{p,n-1} + v_n,\ v_n \sim \mathcal{N}(0, \Sigma_p)$
end for
end for
return Final particle populations, $\{x_{T,p},\ p = 1, \ldots, P\}$ and $\{\theta_{T-1,p},\ p = 1, \ldots, P\}$.

Adaptive MPC: *in silico* performance





$A_d(3,1) = 0.02$ (model mismatch)

At $t = 280$: b_r doubled, d_p increased 1.5-fold

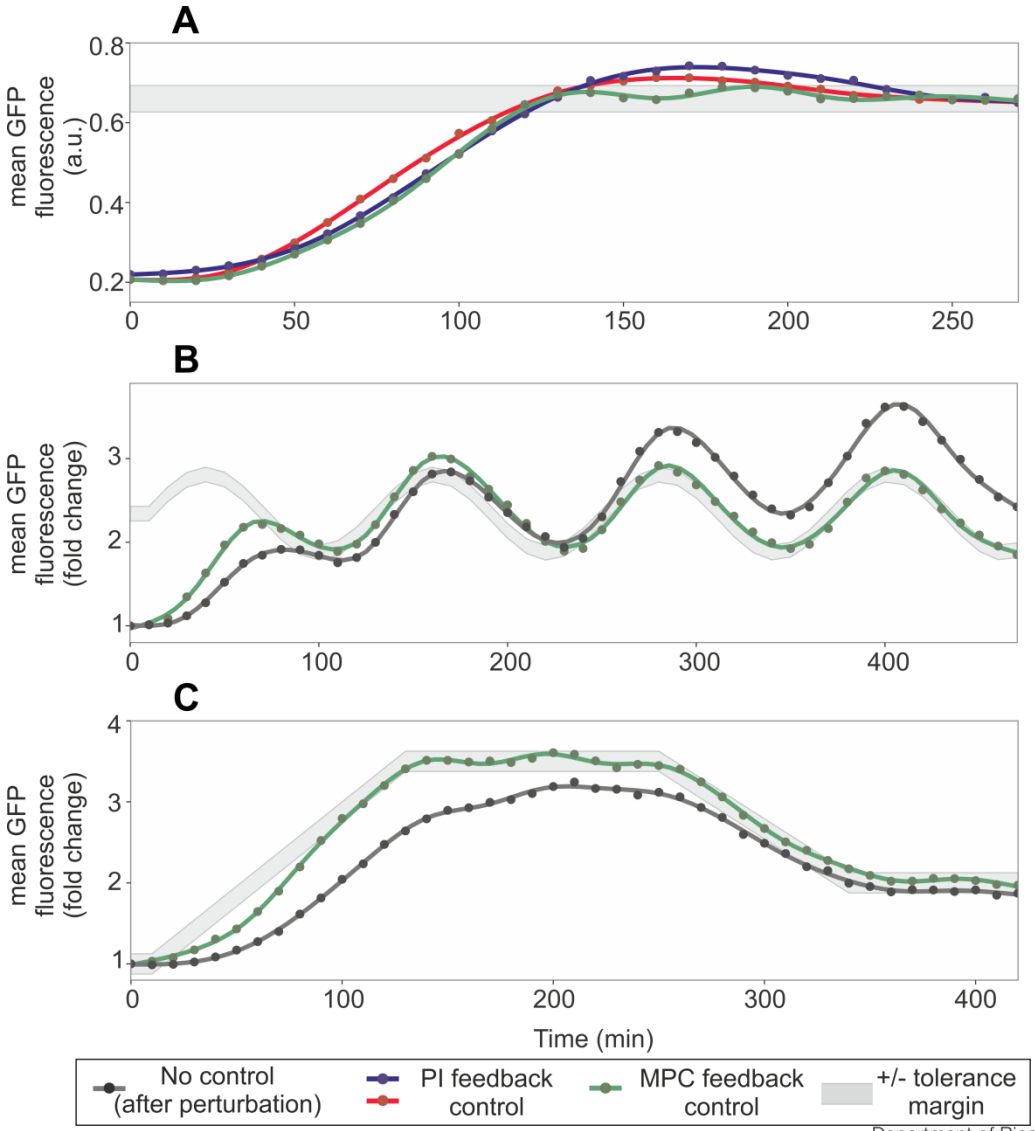
Outline

1. Preliminaries: molecular biology
2. Preliminaries: optogenetics
3. Motivation
4. Experimental system + setup
5. Controllers
6. Experimental results: tracking + disturbance rejection
7. Discussion

Tracking results

Different PI curves
→ different days

Grey lines and dots: application of the same input on a different day

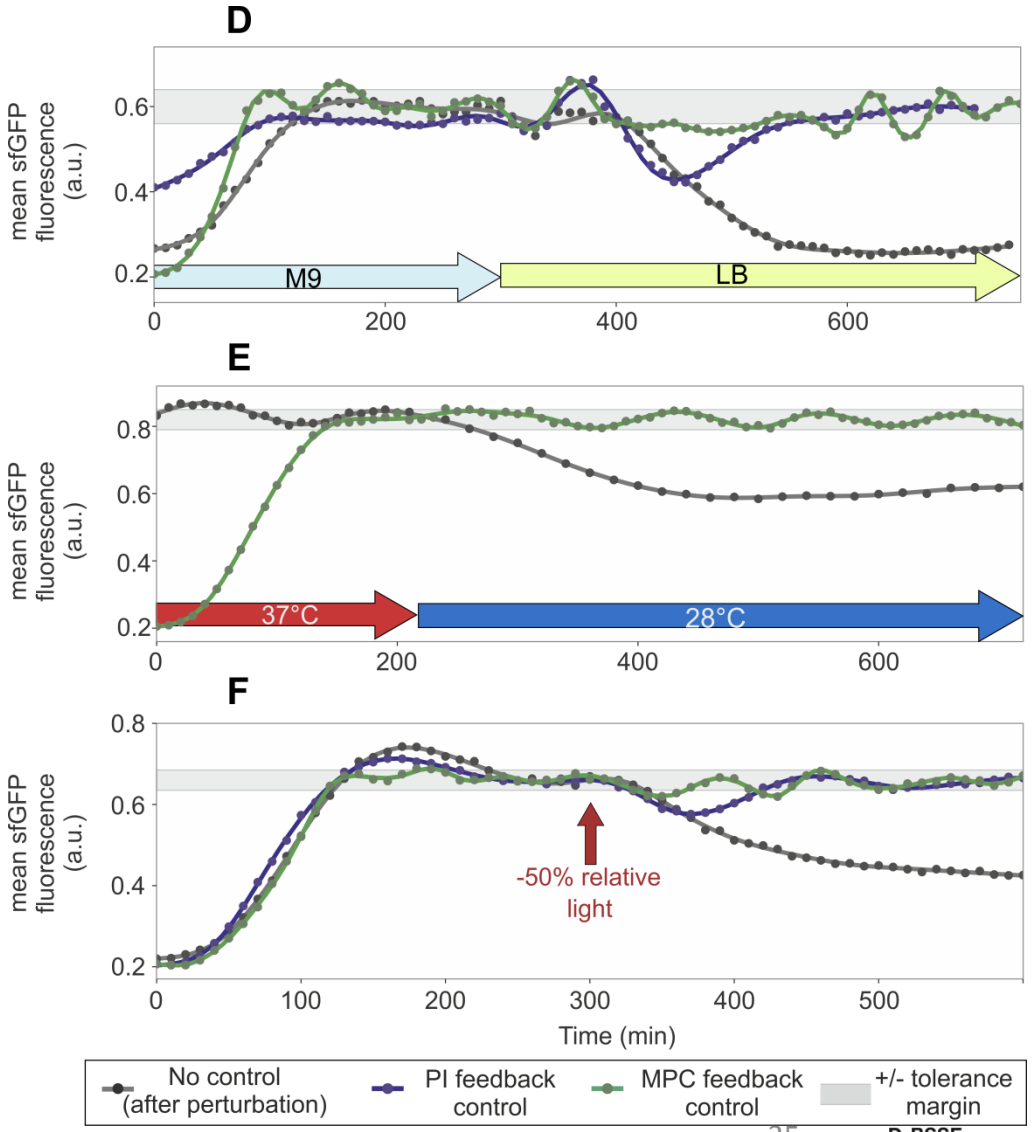


Disturbance rejection

Change of culture medium

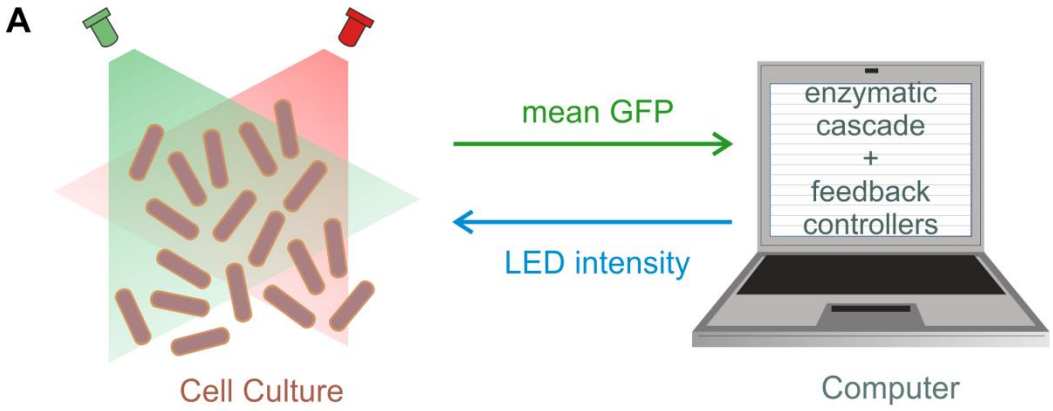
Temperature shift

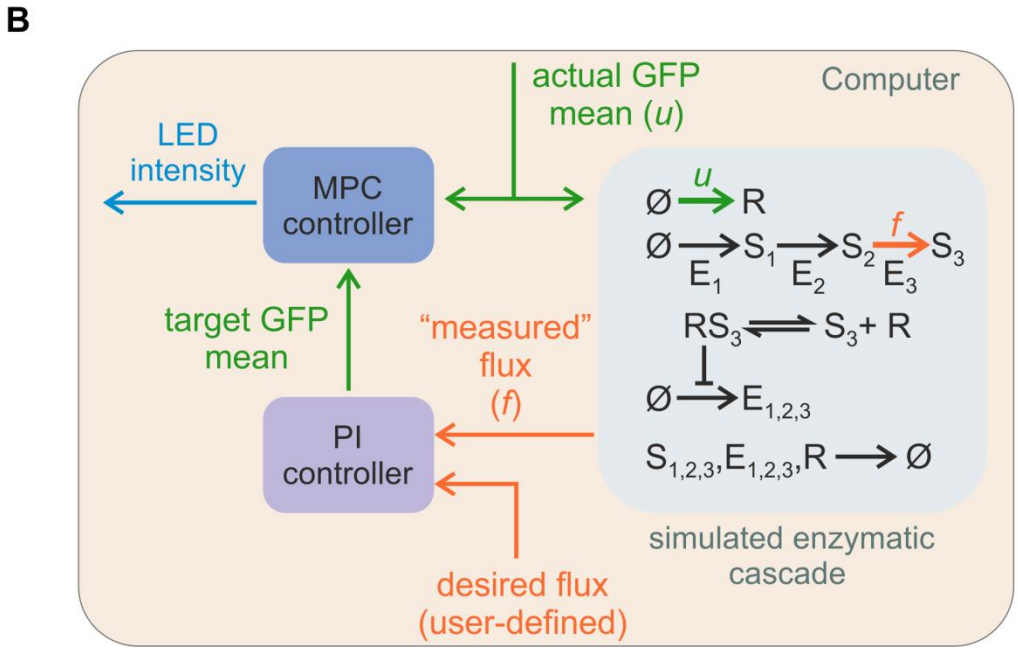
Light input perturbation ("LED damage")



ETH zürich
D-BSSE
Department of Biosystems Science and Engineering

… and some more crazy stuff

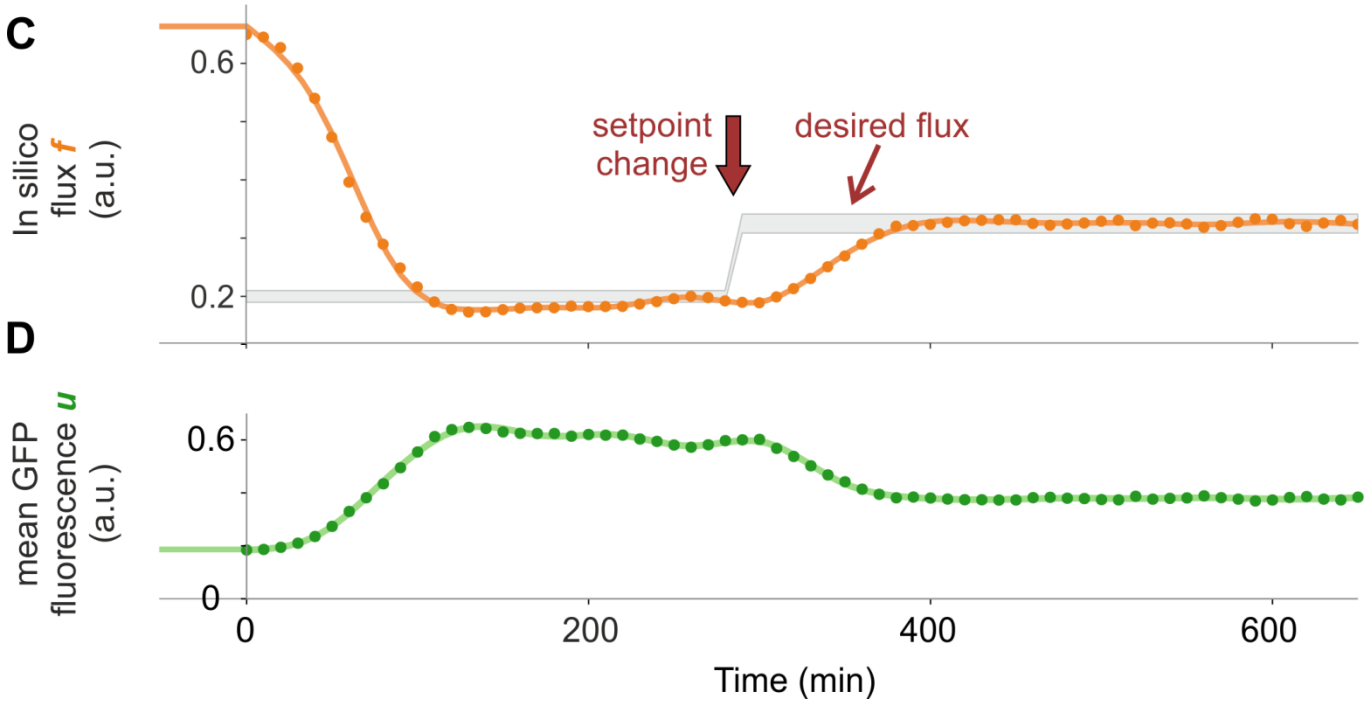




Tracking a given flux profile

Simulated system output…

… controlled by measured GFP mean



Outline

1. Preliminaries: molecular biology
2. Preliminaries: optogenetics
3. Motivation
4. Experimental system + setup
5. Controllers
6. Experimental results: tracking + disturbance rejection
7. Discussion

Discussion

- Integrated framework for automatic optogenetic feedback of liquid cell cultures comprising tailor-made hardware and software
- Demonstrated the strength and flexibility of feedback regulation in comparison to open-loop approaches
- Further exploration of accuracy/complexity tradeoffs alternative control schemes (experimental testbed)
- Experimental platform can be used to generate targeted perturbations for the characterization of intracellular pathways

Possible expansions/applications

- Parallel continuous cell cultures (in progress)
- Multiplexed sampling (e.g. with robotic arm – in progress)c
- More complex control with multiple optogenetic systems in the same cell
- Rich field of applications: bioprocess control, to improve productivity, robustness and batch-to-batch variability
- Online monitoring and optogenetic feedback control of cells inside bioreactors