Scaling up genetic circuit design for cellular computing: advances and prospects

Abhinav Mishra
IIIT-H

INTRODUCTION

Survive and Reproduce <= Signal sensing

Synthetic Biology

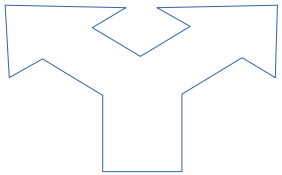
The design and engineering of biologically-based parts, novel devices and system as well as the redesign of existing natural biological system.

(No Universal Definition!)

Key concepts

- Standardization
- Modularity
- Characterization
- Orthogonality

Digital **Analog**



Genetic Circuits

Design

Test

Build

Learn

FOCUS

- **Digital-like biological parts**
- **Tools & Challenges of construction of large scale genetic circuits**
- **Common methods enabling control of genetic Circuits**
- **Automated gene circuits – Design & Software**
- **Obstacles to Modularity + Context effects + metabolic burden**

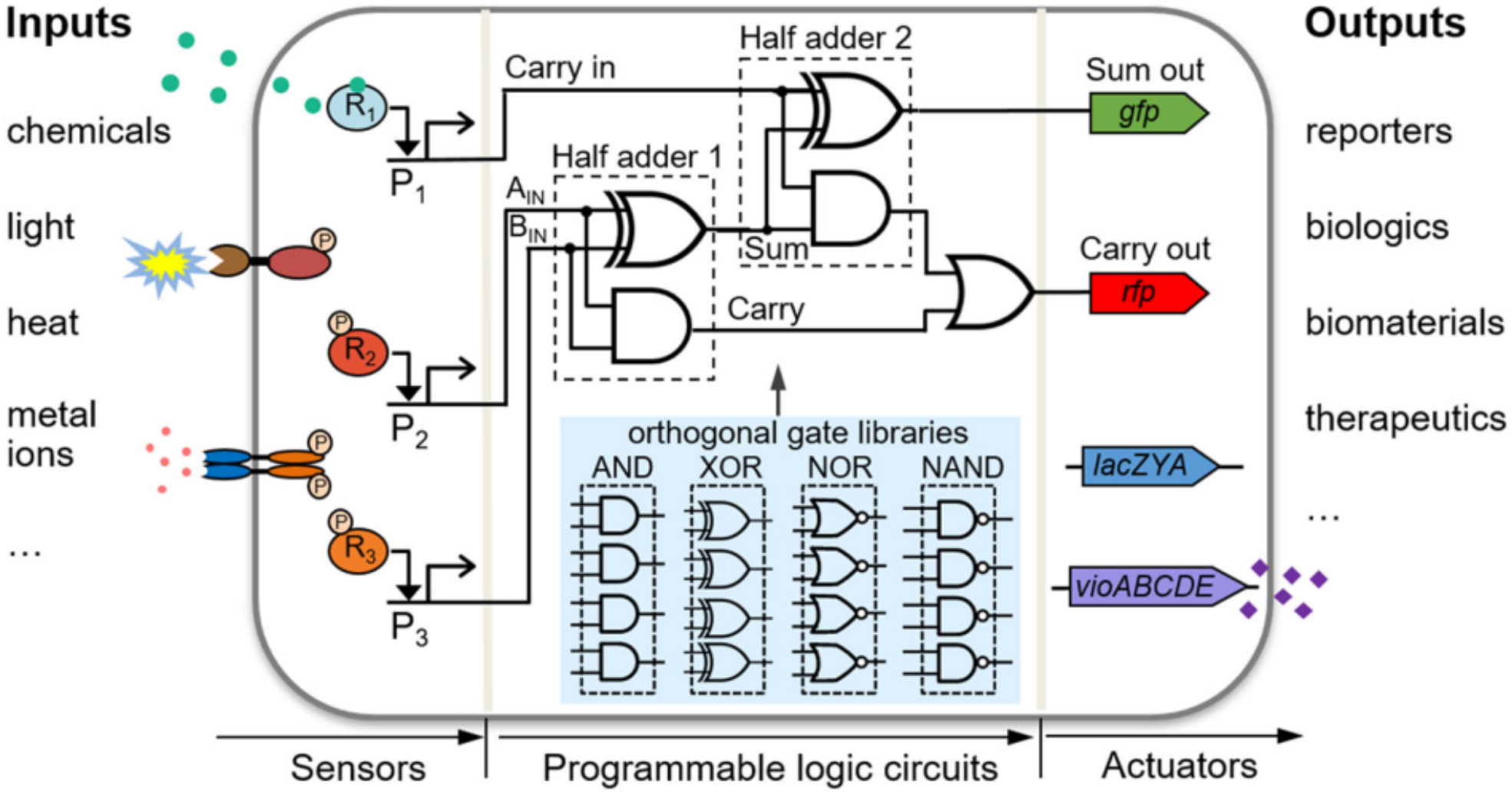
DBTL

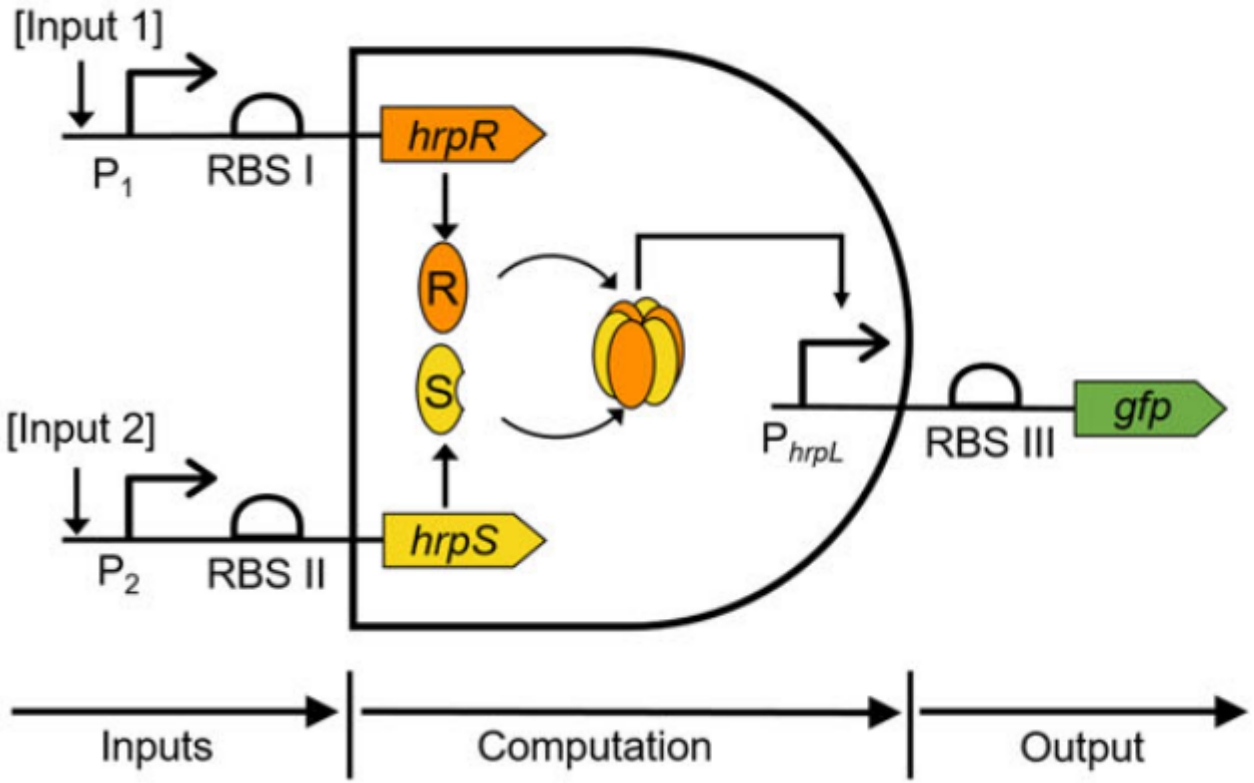
1.Design – arrangement of reusable components to produce biological programs

2.Build – large scale DNA assembly

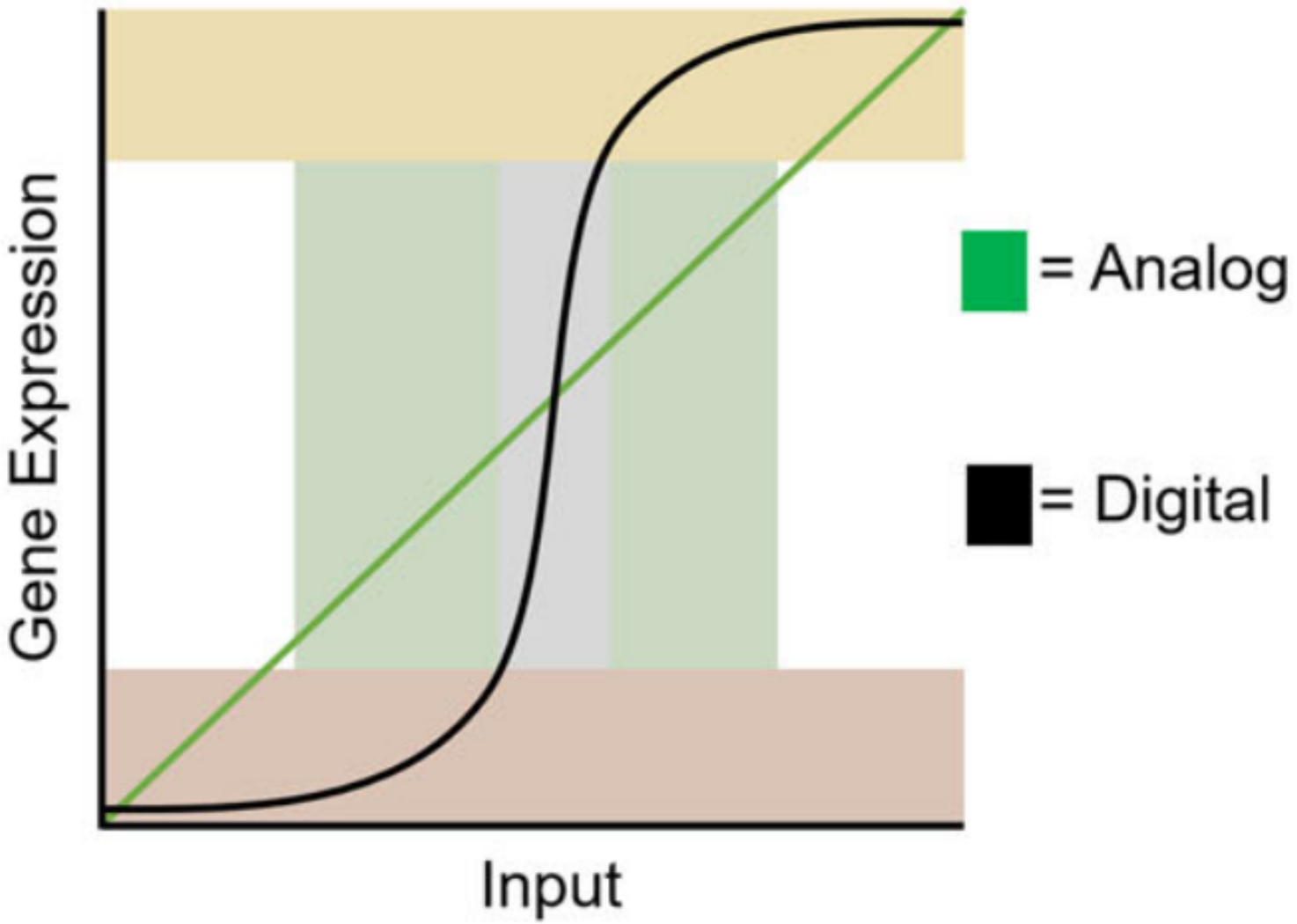
3.Test – High-throughput characterization and debugging tools

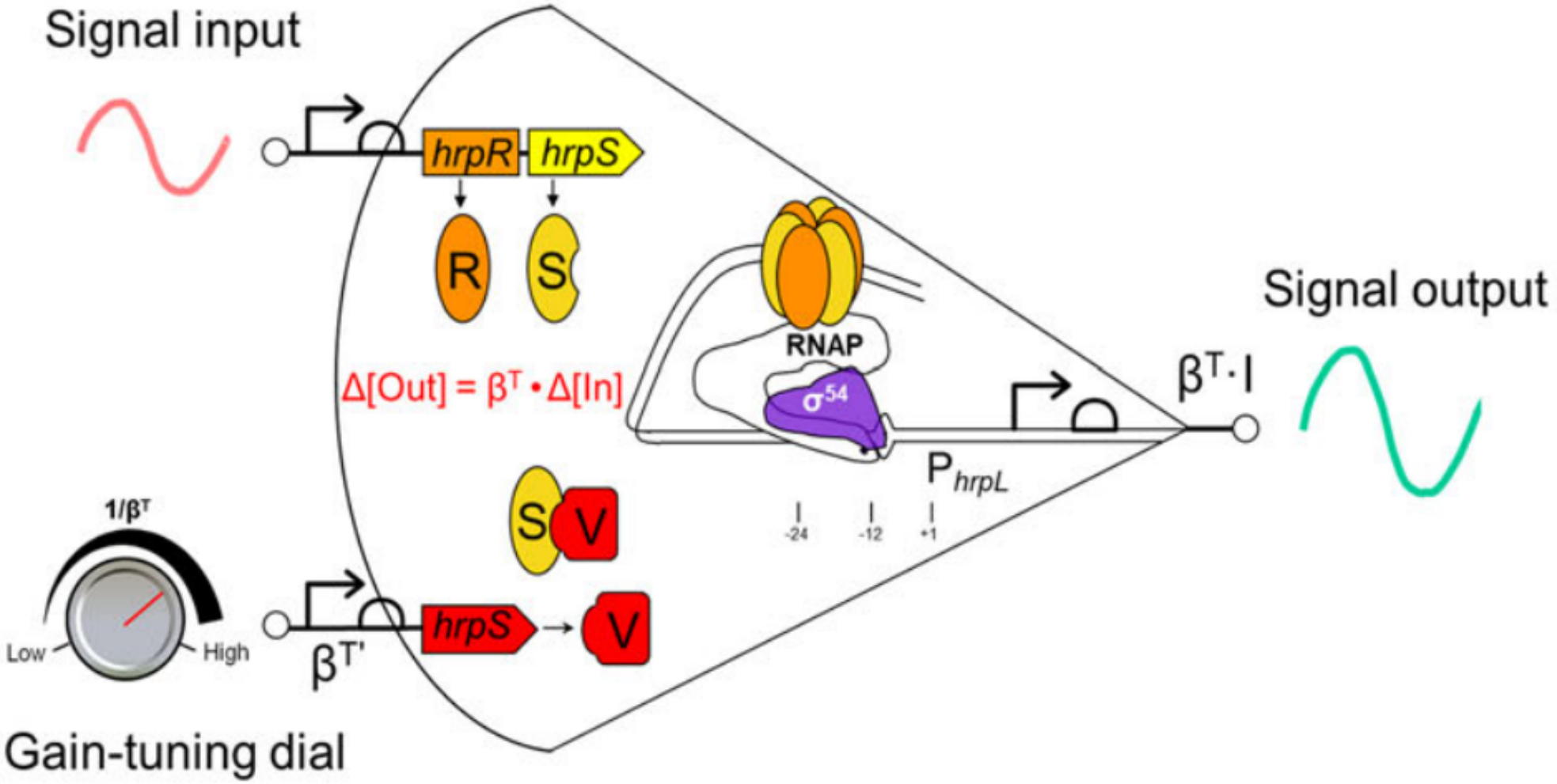
4.Learn – Modelling and circuit design automation


Inputs
chemicals
light
heat
metal ions
…
R1
P1
R2
P2
R3
P3
Carry in
Half adder 2
Half adder 1
AIN
BIN
Sum
Carry
orthogonal gate libraries
AND
XOR
NOR
NAND
Sum out
gfp
Carry out
rfp
lacZYA
vioABCDE
Outputs
reporters
biologics
biomaterials
therapeutics
…
Sensors
Programmable logic circuits
Actuators

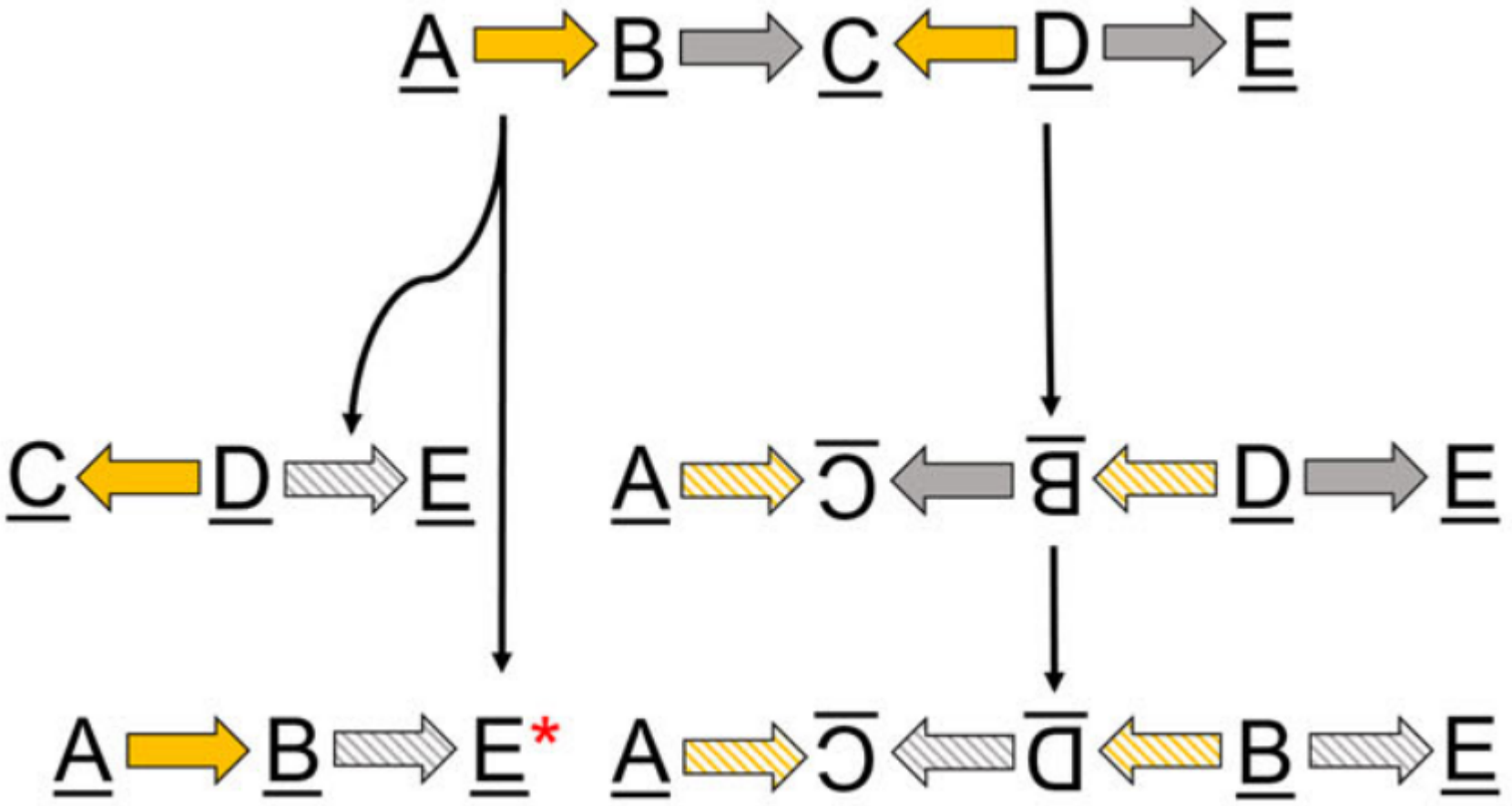


Input 1	Input 2	Output
0	0	0
0	1	0
1	0	0
1	1	1


Gene Expression
Input
= Analog
= Digital

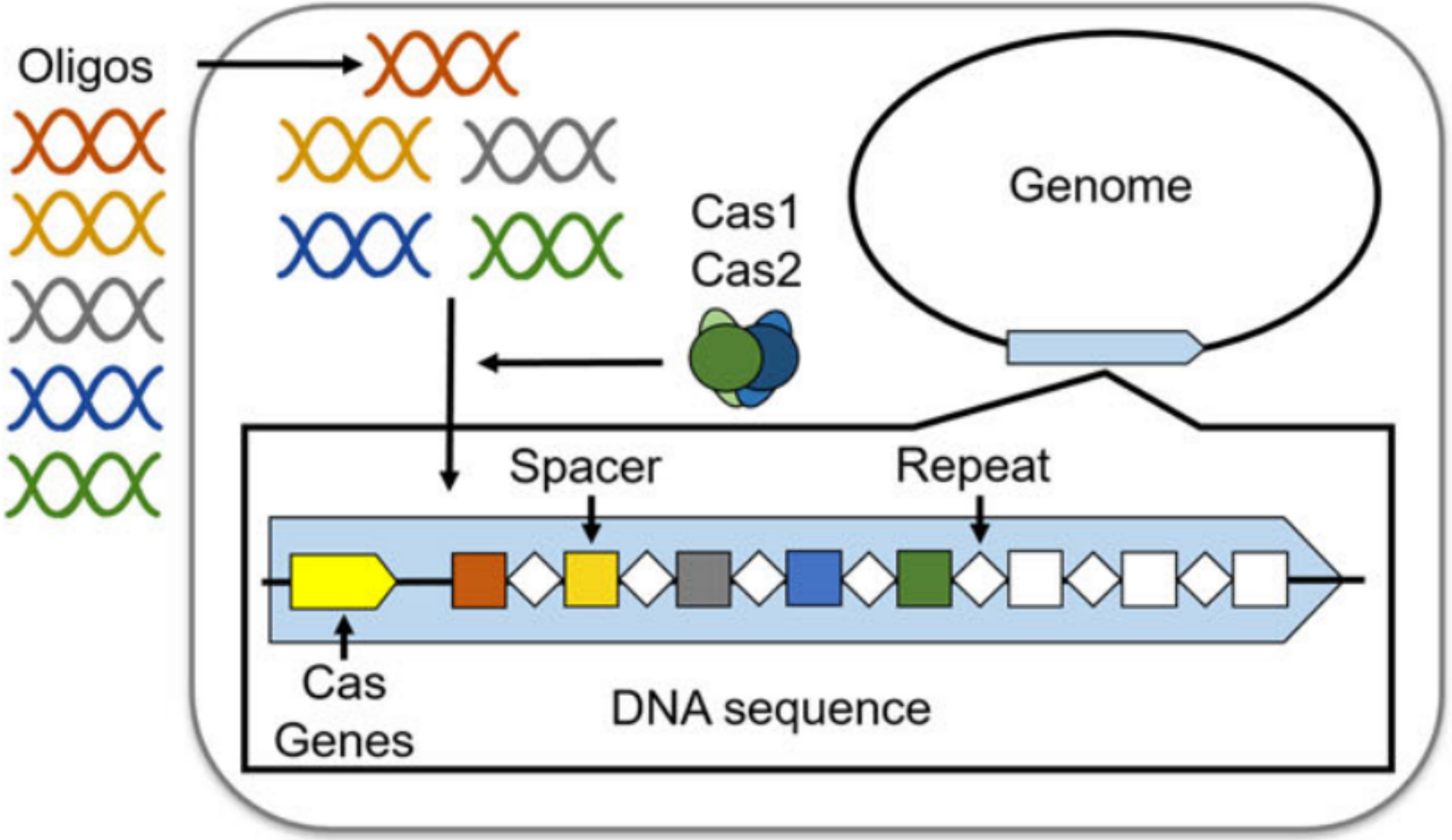

Signal input
hrpR
hrpS
R
S
RNAP
σ54
Signal output
Δ[Out] = βT • Δ[In]
βT·I
PhrpL
-24
-12
+1
S
V
1/βT
Low
High
hrpS
V
βT'
Gain-tuning dial

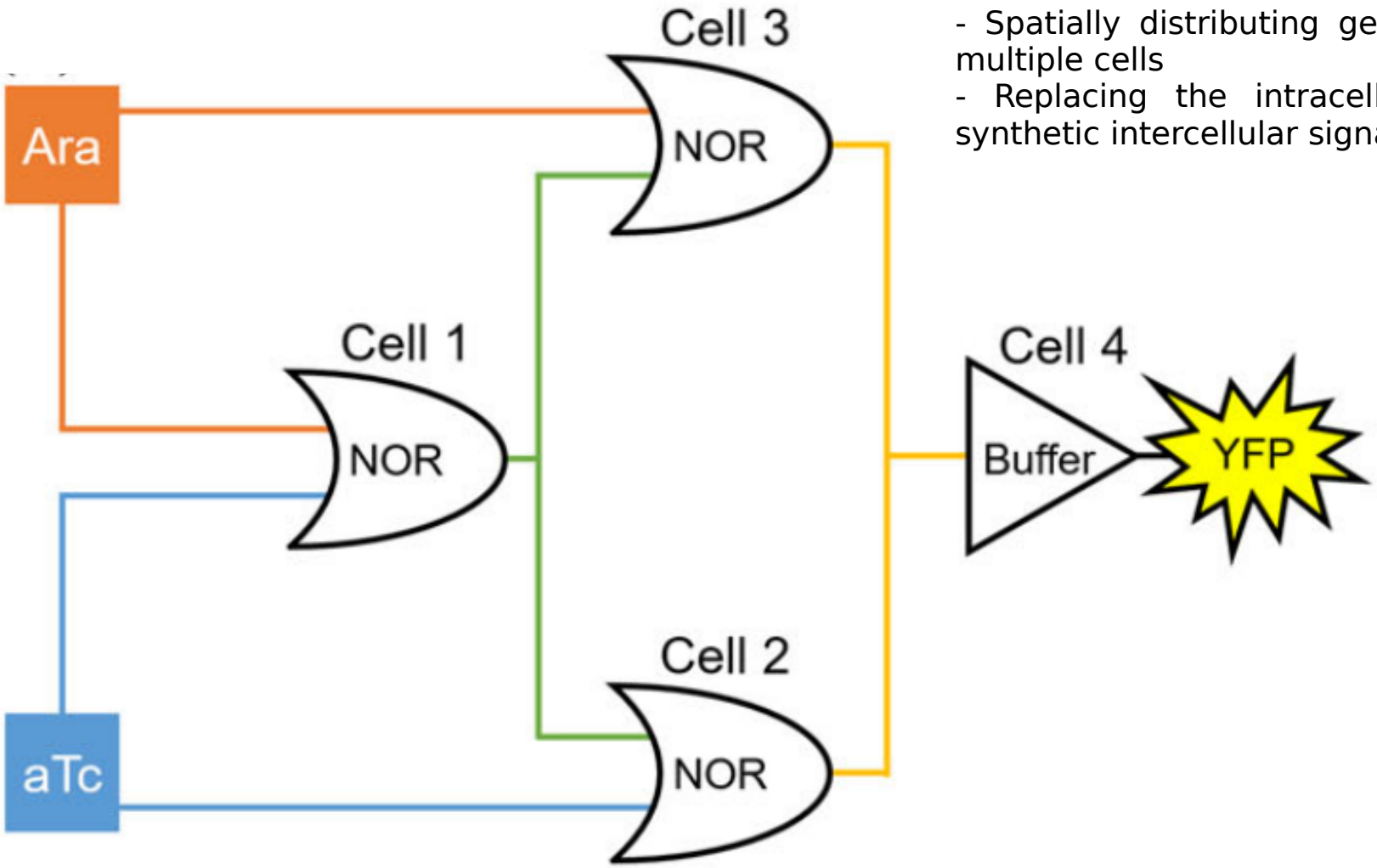
Basic 3- state version



CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR associated protein (Cas) system stores DNA from invading species in the genome to generate an immune response

- Memory can also be stored by assigning colors





- Spatially distributing gene circuits across multiple cells
- Replacing the intracellular ‘wires’ with synthetic intercellular signalling pathways

DESIGN

Control of expression using RNA:

cleavage

pair binding

secondary structure

Expansion of parts (Tools) :

genomic part mining

bioinformatics and site directed mutagenesis

Inducible parts **---> External Control**

customizable DNA sequence binding **--->** **number of orthogonal parts**

PPI **---> multi-input genetic logic gates**

Tuning ‘knobs’

To adjust the dynamic properties of a system to respond effectively to expected inputs and produce desired outputs depending on the need.

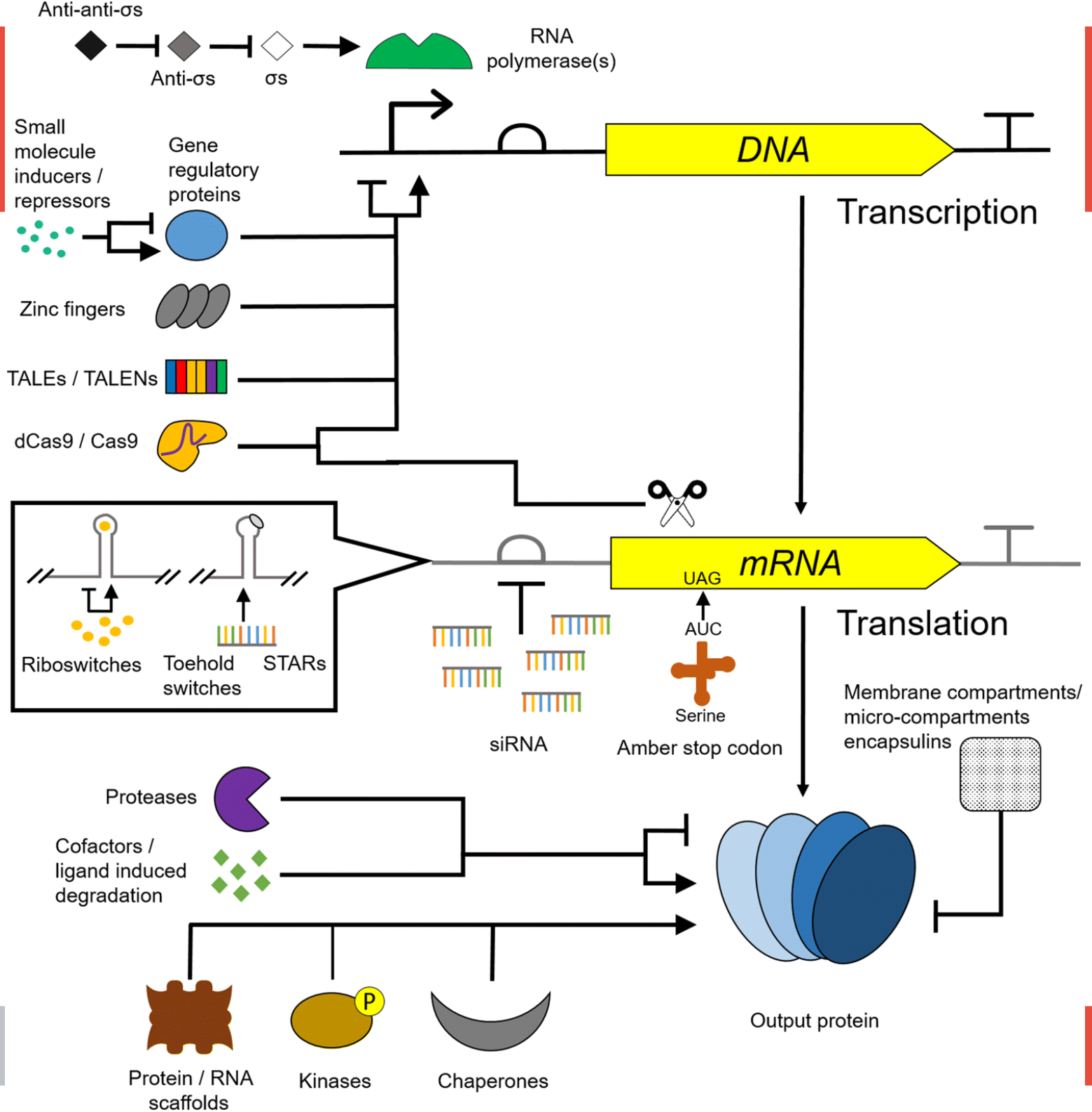
- strength of the promoter sequence
- hybrid combinations of promoter sequences
- operator site modification
- ribosome binding site (RBS) modification
- altering plasmid copy number
- using decoy DNA operators
- RNA interference (RNAi)
- degradation tags or co-expression with sequestering proteins or molecules

Improvement

Positive feedback loops and signal cascades have improved the ON/OFF ratio.

"Leaky" gene expression

Cooperativity has been improved using oligomerization domains

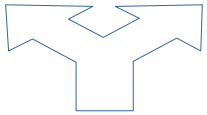

Anti-anti-σs
Anti-σs
σs
RNA polymerase(s)
Small molecule inducers / repressors
Gene regulatory proteins
DNA
Transcription
Zinc fingers
TALEs / TALENs
dCas9 / Cas9
Riboswitches
Toehold switches
STARs
mRNA
UAG
AUC
Translation
siRNA
Serine
Amber stop codon
Membrane compartments/ micro-compartments encapsulins
Proteases
Cofactors / ligand induced degradation
P
Protein / RNA scaffolds
Kinases
Chaperones
Output protein

BUILD

Chemically synthesized genes => Expensive + No standardization

Initial (2008) --> lengthy stepwise restriction enzyme mechanisms (eg. BioBricks™)

MoClo **Goldenbraid**



Latest --> “One Pot” ;multiple fragments can be assembled at once in a defined order

DNA assembly protocols

Gibson assembly (non-synthesis based method)

homology of overlapping single stranded DNA which also avoids the necessity of removing forbidden sequences (such as restriction enzyme sites) in the sequences.

MODAL (Modular Overlap-Directed Assembly with Linkers)

add modular prefixes and suffixes.

BASIC (Biopart Assembly Standard for Idempotent Cloning)

exploits orthogonal linkers to avoid PCR entirely and achieved over 90% accuracy with a 7-part reaction.

Golden Gate-Gibson (3G)

combines overhang assembly with Golden Gate style part libraries.

TEST

FRET < < FACS ⇐ multiple cellular properties in each cell (size, granularity and fluorescence analysis)

⇑ Localization and characterization

fluorescence measure of a regulatory sequence (coding) (3' end) ≃ measuring the final product of gene expression

Secondary structures ⇕

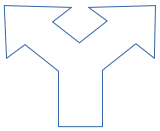
High Copy Number ⇓ RNA-seq ⇒ antisense promoters terminator malfunction media related failure

Metabolic load ⇒ **Reference Promoter Unit (RPU)**

Var < 4%

Insulators (Ribozymes that cleave 5' UTR)

Direct Sampling
(HPLC, MS)



Real-time dynamics
(RNA aptamers, FRET probes)

Drawbacks:

lack of shared cellular resources

imbalance in ΔP

Cell free in vitro systems (Prototyping)

e.g.

E. coli transcription-translation based cell free system (TX-TL) for prototyping promoters and negative feedback loops

Microfluidics - precise manipulation of small amounts of fluids in the micro and nanoliter scale.
Continuous - using oil and water to generate a controllable liquid stream
Digital - using voltage to control the movement of individual droplets on a conductive material
Advantages - artificial cell-free entities prototyping in parallel

FADS (fluorescent activated droplet sorting)

LEARN

Tools :

Genocad

CellDesigner

Biojade

SynbioSS

Tinkercell

Visual GEC

Cello

SBOL (Synthetic Biology Open Language)

SBML (Systems Biology Markup Language)

Model Analysis in general platfroms
(MATLAB & COPASI)

Cello

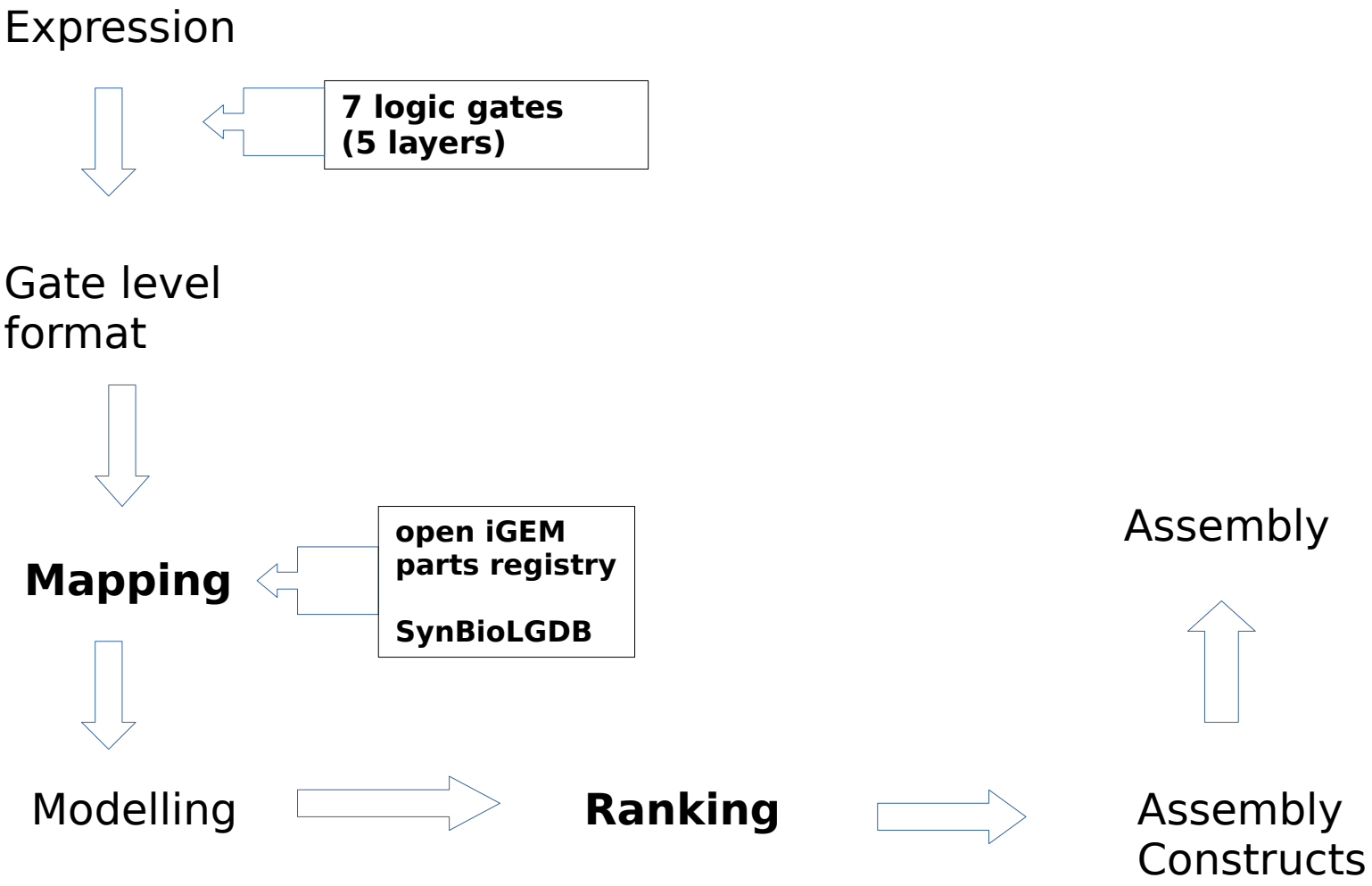
Verilog logic programming language => circuit function

User constraints => parts and organisms (searchable design space)

Circuits are modified to be *compatible* with a library containing *NOR and NOT gates based on repression*.

System is simulated to **predict circuit performance** (*Parameters : load, population variability, growth, and connectivity in terms of RNA polymerase (RNAP) flux*)

GENE CIRCUIT DESIGN AUTOMATION FLOW



1-bit full subtractor circuit design – function input

Inputs			Outputs	
X	Y	B_{in}	Diff	B_{out}
0	0	0	0	0
0	0	1	1	1
0	1	0	1	1
0	1	1	0	1
1	0	0	1	0
1	0	1	0	0
1	1	0	0	0
1	1	1	1	1

Boolean logic expression (SOP)

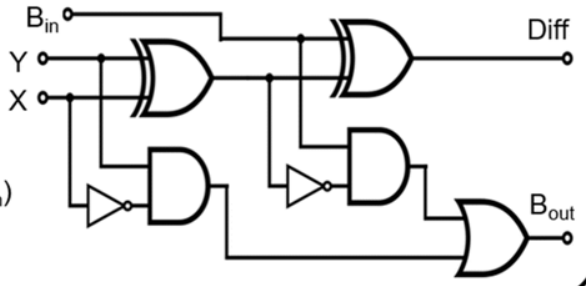
$Diff = X'Y'(B_{in}) + X'Y(B_{in})' + XY'(B_{in})' + XY(B_{in})$

$B_{in} = X'Y'(B_{in}) + X'Y(B_{in})' + X'Y(B_{in})' + XY(B_{in})$

Optimization and logic synthesis

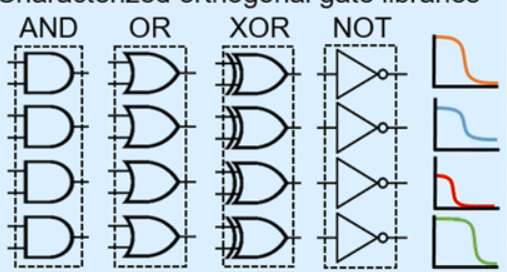
$Diff = X \oplus Y \oplus B_{in}$

$B_{out} = X'(B_{in}) + X'Y + Y(B_{in})$



Technology mapping and function simulation

Characterized orthogonal gate libraries



Boolean logic model

Differential equation models

…

Ranked mapped gene circuit candidates

Toxicity	Dynamic range	Ranking
2%	500	1
8%	400	2
6%	220	3
…	…	…

Challenges in Automated Software Design

Mathematical models that are *predictive of circuit function*

- **nonlinear features** of the biochemical interactions (given circuit)

Mechanistically : specific interactions of elements (DNA/Protein) of the cell **should be known**.

Chemical reactions are *discrete* and *stochastic* (Scope of the answer!)

Still, ODEs, PDEs, SDEs, Sampling/Integration of CME, etc.

Methods for inferring parameters for models of specific circuits is just as important.

Advanced paradigms in cellular computing

Sequential Logic over Combinatorial Logic

State machines can be one of a number of finite states at any given time, with access to states dependent on predetermined sequence of events triggered by various conditions - most complex reported has 16 different positions

Distributed Computing – avoids any potential “CrossTalk”

Most complex system to date : a 1-bit full binary adder; incorporating 22 separate gates distributed amongst 9 specialized mammalian cell types in a complex three dimensional environment

Cell–cell signalling to induce structural self-organisation of tissues and pattern formation

Concluding Remarks

Applications :

biomanufacturing

biosensing

Biotherapy

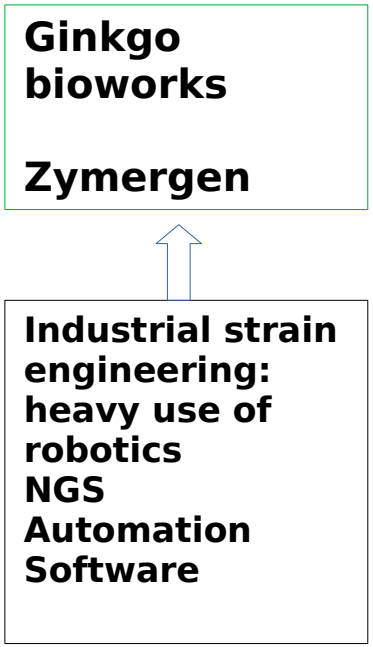
Problems to overcome :

significant lack of predictability

modularity and standardization

fundamental gaps in knowledge of basic biological processes

gap between proof of concept and industrially sized production



Reference

Xiang, Y., Dalchau, N. & Wang, B. Nat Comput (2018) 17: 833. https://doi.org/10.1007/s11047-018-9715-9