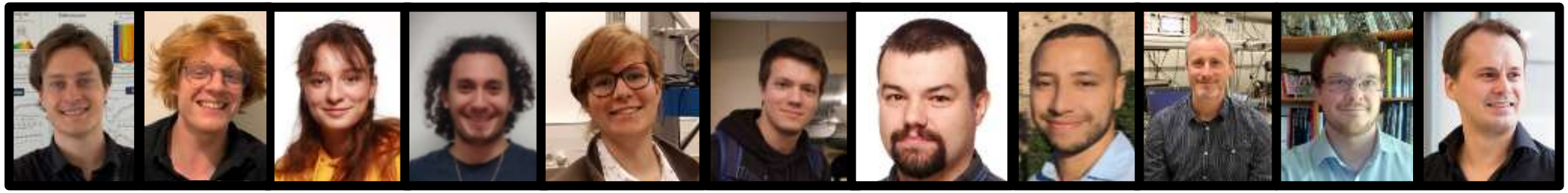
# Machine learning in nonlinear and ultrafast photonics




**John M. Dudley**

Andrei V. Ermolaev, Lilian Emonin, Mathilde Hary, Mehdi Mabed, Coraline Lapre, Lev Leybov,
Piotr Ryczkowski, Anas Skalli, Daniel Brunner, Christophe Finot, Goery Genty

# Machine learning is impacting all of science

The tools of artificial intelligence and machine learning are having tremendous impact across science, including of course in ultrafast and nonlinear optics and photonics

## Nobel Prize in Physics 2024

John J. Hopfield



Geoffrey Hinton

The Nobel Prize in Physics 2024 was awarded jointly to John J. Hopfield and Geoffrey Hinton "for foundational discoveries and inventions that enable machine learning with artificial neural networks"

They used physics to find patterns in information

## Nobel Prize in Chemistry 2024

David Baker

Demis Hassabis

John Jumper

The Nobel Prize in Chemistry 2024 was divided, one half awarded to David Baker "for computational protein design", the other half jointly to Demis Hassabis and John Jumper "for protein structure prediction"

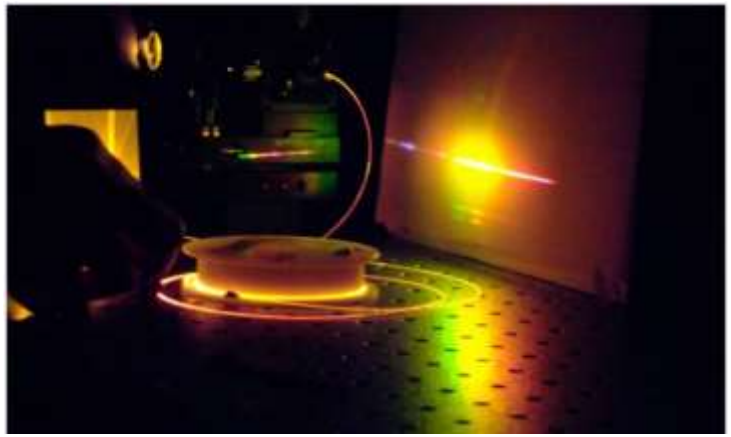
# Why should we care about ultrafast photonics?

Several Nobel prizes have been associated with ultrafast photonics

1999 Femtochemistry
2005 Frequency comb
2009 Fibres for communications
2018 Chirped pulse amplification
2023 Attoscience

Ultrafast laser sources based on soliton shaping
Fibre supercontinuum generation
Linear guidance properties of fibre waveguides
Chirp control in optical fiber propagation
Ultrafast laser sources based on soliton shaping

At high intensities and with short pulses, dispersive and nonlinear interactions in waveguides create a vast landscape of novel propagation phenomena that reveal fascinating physics

Localised solitons, frequency conversion, broadband supercontinuum generation …

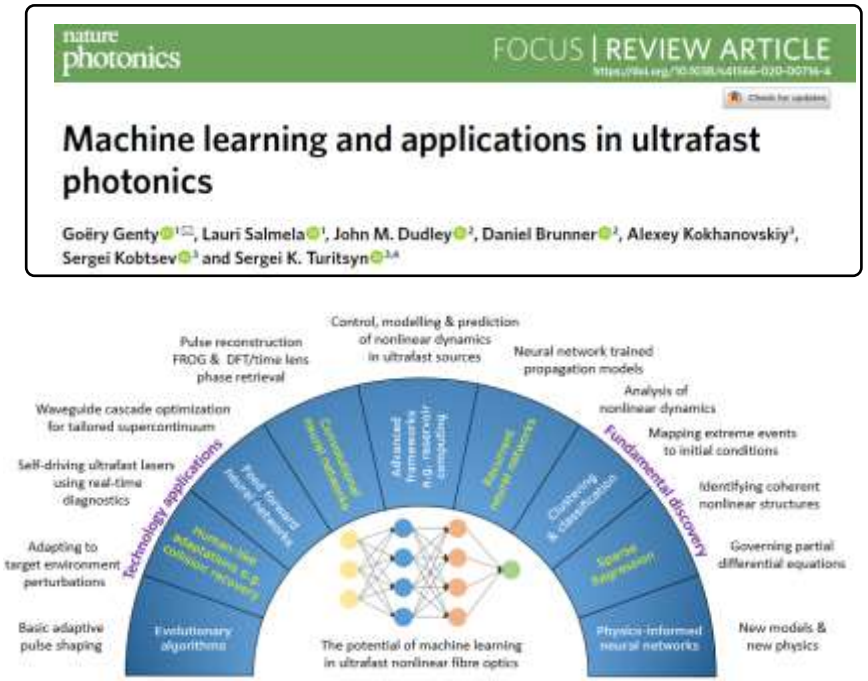
J. M. Dudley *et al*. *EPL* Perspective **151** 55001 (2025)

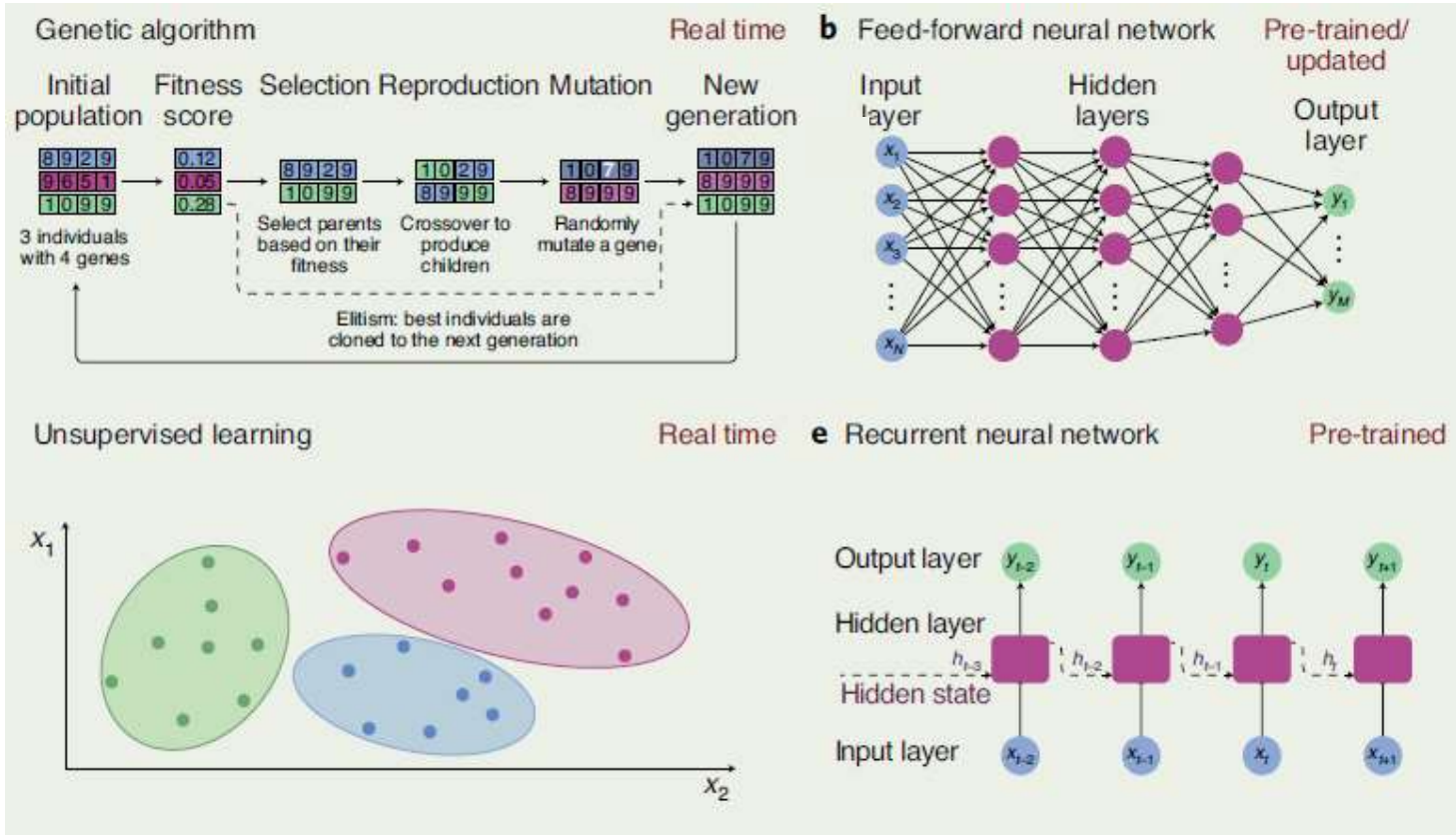
# Questions to be answered during this talk

How machine learning can help us better understand or analyse ultrafast nonlinear fibre optics

How we can use machine learning methods to optimise fibre based sources

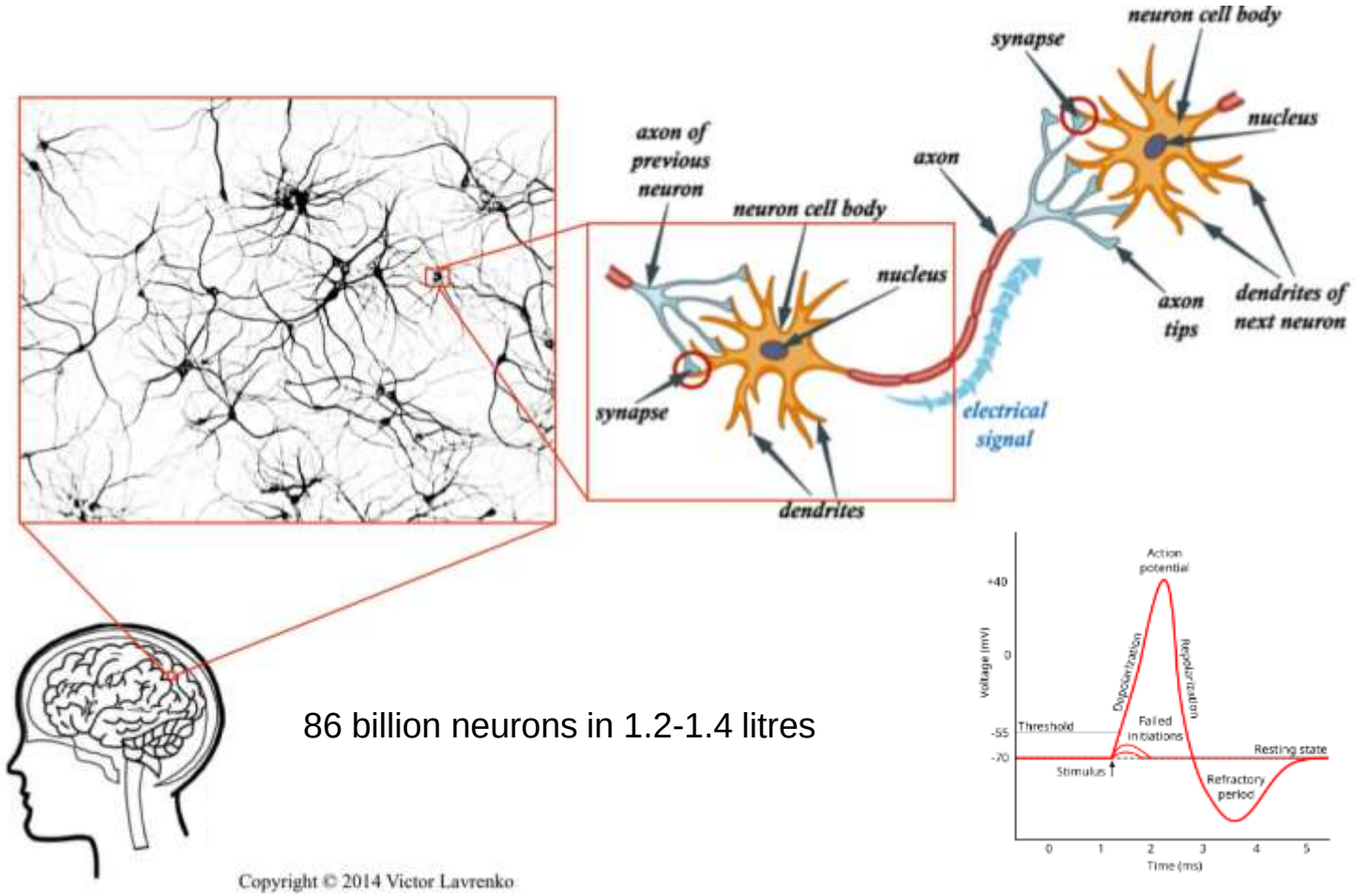
How we can exploit nonlinear and dispersive interactions in fibre for computation



# Neural network basics

An artificial neuron is an abstraction of a biological neuron that receives inputs (electrical or chemical), processes these inputs to fire an action potential, then transmits this to other neurons.




Copyright © 2014 Victor Lavrenko

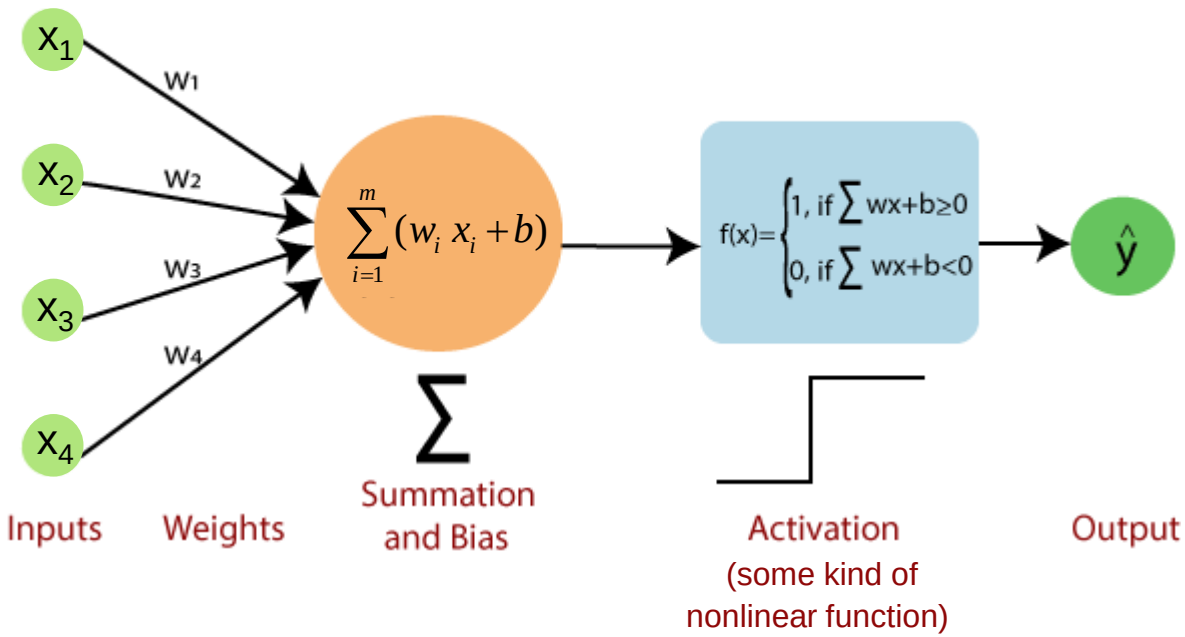
**A neuron is a nerve cell** that belongs to the nervous system. Like other cells, it has a **cell membrane, cytoplasm, and nucleus**.

What makes it unique is its ability to **generate and transmit electrical signals**.

Its specialized structures (**dendrites and axon**) allow communication with other cells

# Neural network basics

An artificial neuron is an abstraction of a biological neuron that receives inputs (electrical or chemical), processes these inputs to fire an action potential, then transmits this to other neurons.




**Receives inputs** (numbers representing **features**, e.g., pixel values, voltage, intensity …)

**Multiplies each input by a weight** (how important that input is)

**Adds them together** with an offset term (bias)

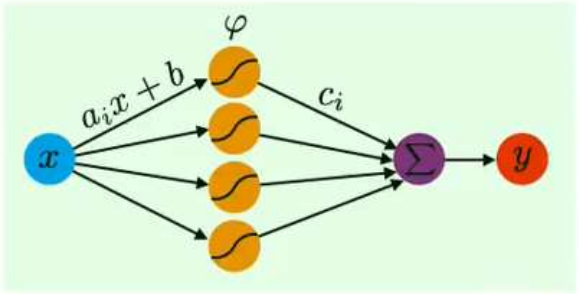
**Applies an activation function** (typically a step function) to decide the output.

1943: Warren McCulloch and Walter Pitts
1958: Frank Rosenblatt
1989: George Cybenko

towardsdatascience.com
javatpoint.com

# The Universal Approximation Theorem (Cybenko)

A single layer of neurons can approximate an arbitrary continuous function given enough neurons.




1 hidden layer perceptron:

$$y \approx f(x,a,b,c) \overset{\text{def.}}{=} \sum_{i=1}^{p} c_i \varphi(a_i x + b_i)$$

Training:

$$\min_{a,c,b} E(a,b,c) \overset{\text{def.}}{=} \frac{1}{2n} \sum_{k=1}^{n} |y_k - f(x_k,a,b,c)|^2$$

$$\mathcal{L} = \frac{1}{2n} \sum_{k=1}^{n} \left(y_{\text{guess}}^{(k)} - y_{\text{target}}^{(k)}\right)^2 \qquad \frac{\partial \mathcal{L}}{\partial a}, \frac{\partial \mathcal{L}}{\partial b}$$

"Training" involves fitting the network output to some target function via optimization of the **weights and biases** of the neurons, as well as the **output weights**, to minimize some loss function.

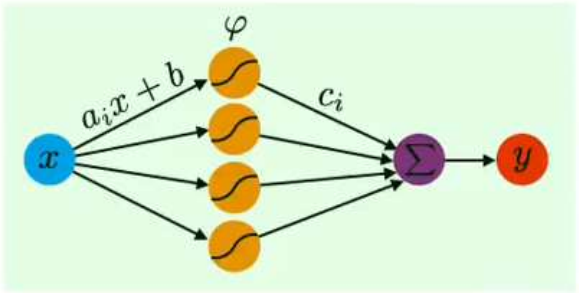
A hidden layer with M neurons and N inputs has M(N+1) trainable parameters.

We minimise a loss function

Gabriel Peyré
@gabrielpeyre

# The Universal Approximation Theorem (Cybenko)

A single layer of neurons can approximate an arbitrary continuous function given enough neurons.

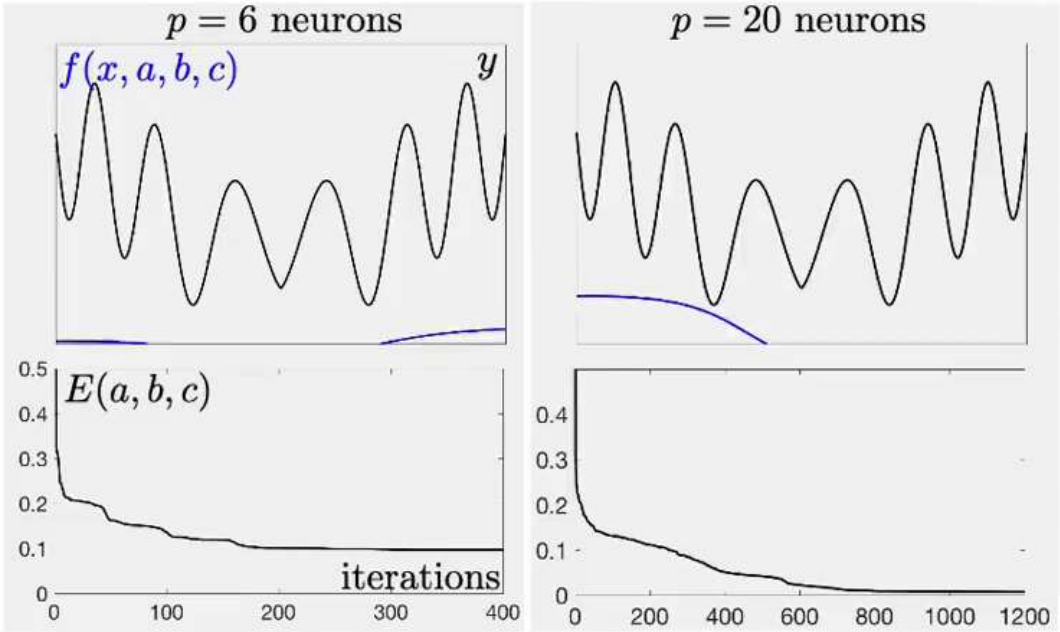



1 hidden layer perceptron:

$$y \approx f(x,a,b,c) \stackrel{\text{def.}}{=} \sum_{i=1}^{p} c_i \varphi(a_i x + b_i)$$

Training:

$$\min_{a,c,b} E(a,b,c) \stackrel{\text{def.}}{=} \frac{1}{2n} \sum_{k=1}^{n} |y_k - f(x_k,a,b,c)|^2$$






"Training" involves fitting the network output to some target via optimization of the **weights and biases** of the neurons, as well as the **output weights**, to minimize a loss function.

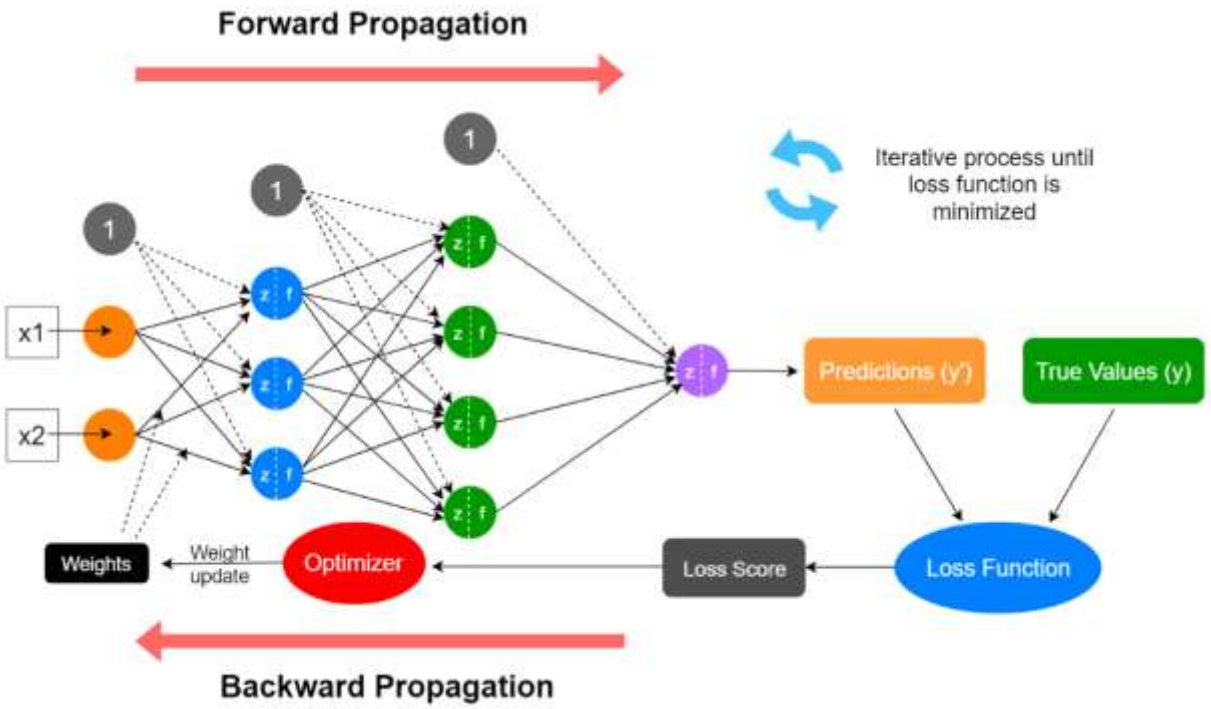
A hidden layer with M neurons and N inputs has M(N+1) trainable parameters.

We minimise a loss function

**Deep learning** just combines multiple layers of neurons

Gabriel Peyré
@gabrielpeyre

# Training and Testing




## Training Phase

The network **learns patterns** from a labeled dataset

Data is passed forward to produce predictions

**Loss function** measures error between prediction and true value (ground truth)

**Backpropagation** adjusts weights to minimize error

Repeated over many **epochs** until performance stabilizes

## Testing Phase

Uses **unseen data** (not used during training)

Evaluates how well the model **generalizes**

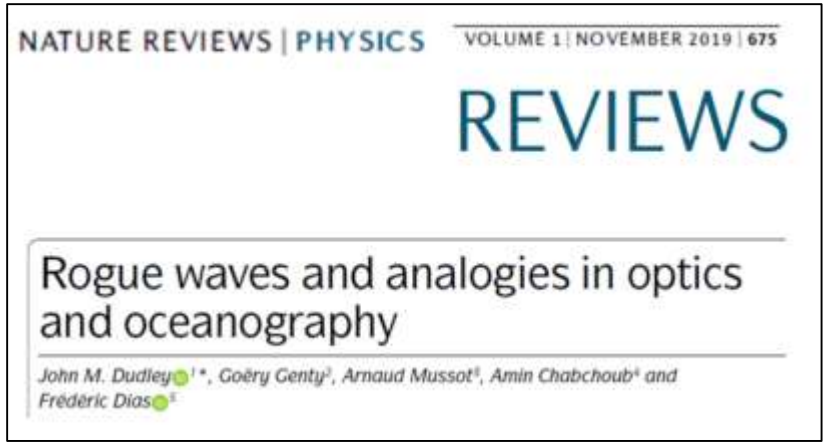
No weight updates — **forward pass only**

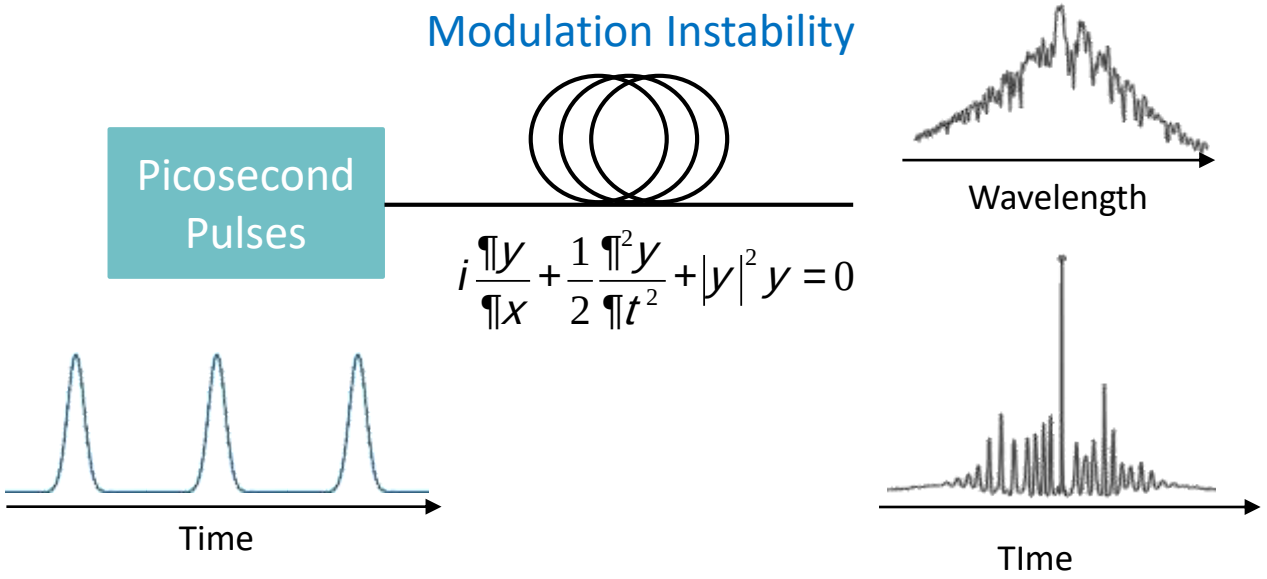
Metrics such as **accuracy, RMS error** are computed

# Neural networks to understand nonlinear propagation in optical fibre

# Noise in highly nonlinear fibre propagation – modulation instability

Modulation instability (MI) is a fundamental property of nonlinear systems where a small perturbation on a continuous wave grows exponentially – observed in water waves, plasmas, optical fibres



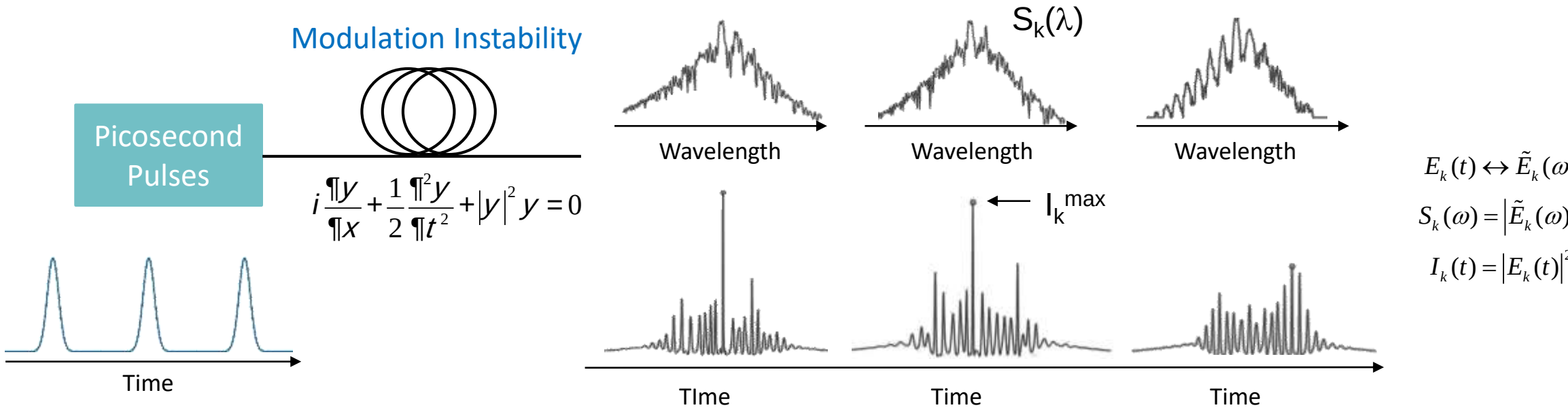




When stimulated by a coherent modulation, MI leads to coherent solitons & “breathers”

When stimulated by noise, MI leads to chaos & “rogue wave” statistics

# Inferring time domain statistics from spectral measurements

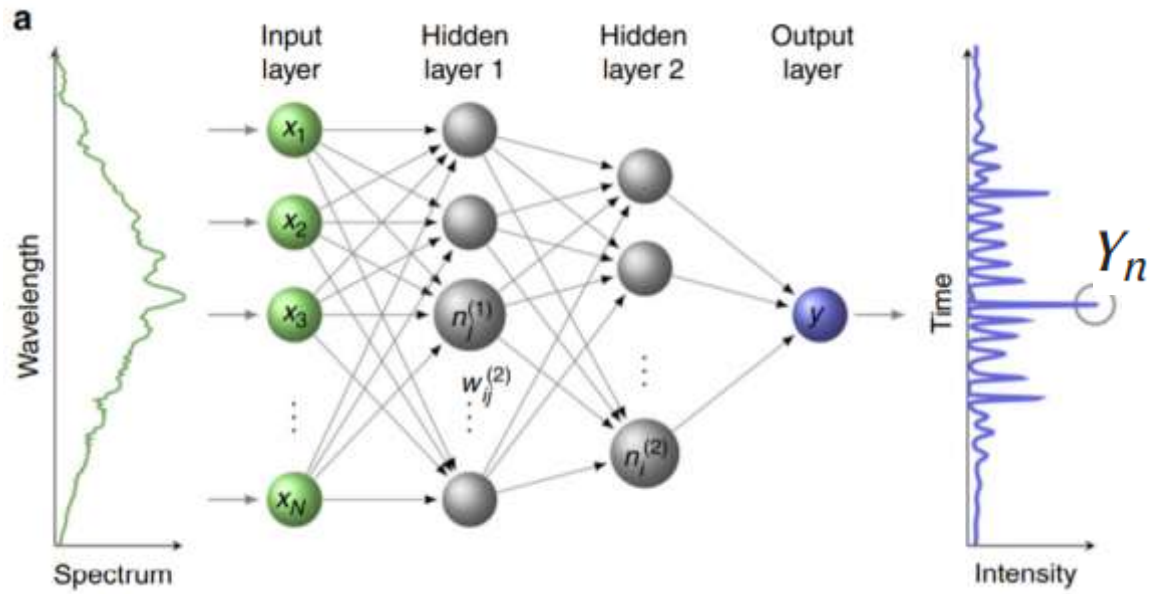
Measuring modulation instability in the time domain is very hard (requires a time lens) but real-time spectral measurements using the dispersive Fourier transform are much easier.




Can we configure a neural network to map shot-to-shot spectra to the peak intensity of the corresponding temporal intensity? That is, can $S_k(\lambda)$ predict $I_k^{max}$ (rogue wave intensity) ?

# Training a neural network to analyse modulation instability

We use a standard neural network architecture and use simulations to train a network to correlate the full MI spectrum with the corresponding peak of the associated temporal intensity profile.




$$\mathbf{X}_n = [x_1, x_2 \ldots x_N]$$

**Architecture**

Feedforward

2 hidden layers (30, 10 nodes)

**Training using NLSE simulations**

$(\mathbf{X}_n, Y_n)$ , $n = 1 \ldots 30{,}000$

$\mathbf{X}_n = [x_1, x_2 \ldots x_N]$ , $N = 121$

300 epochs of 30,000 simulations

# Neural networks can detect extreme temporal events from spectra

Neural networks have been successfully used to correlate the complex spectral and temporal properties of modulation instability and supercontinuum generation, allowing us to build up time-domain statistics from easy to implement real-time spectral intensity measurements

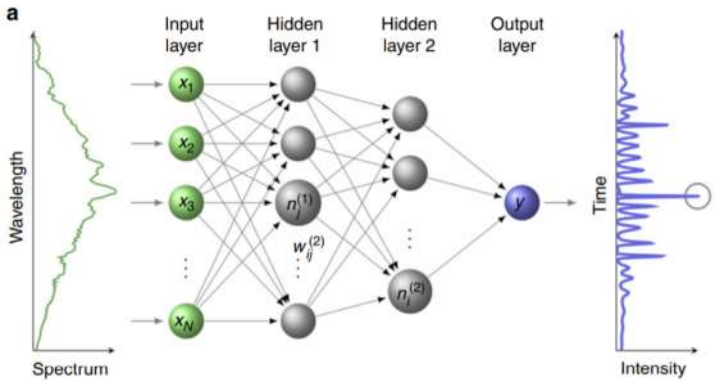
NATURE COMMUNICATIONS | (2018)9:4923

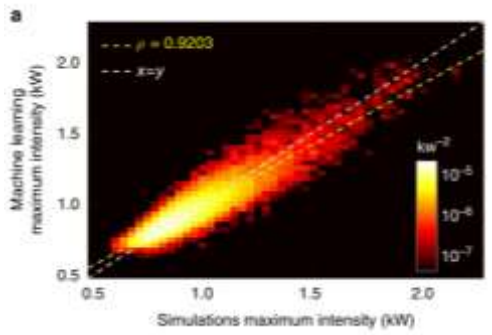
ARTICLE

DOI: 10.1038/s41467-018-07355-y OPEN

## Machine learning analysis of extreme events in optical fibre modulation instability

Mikko Närhi [1], Lauri Salmela [1], Juha Toivonen[1], Cyril Billet[2], John M. Dudley [2] & Goëry Genty[1]

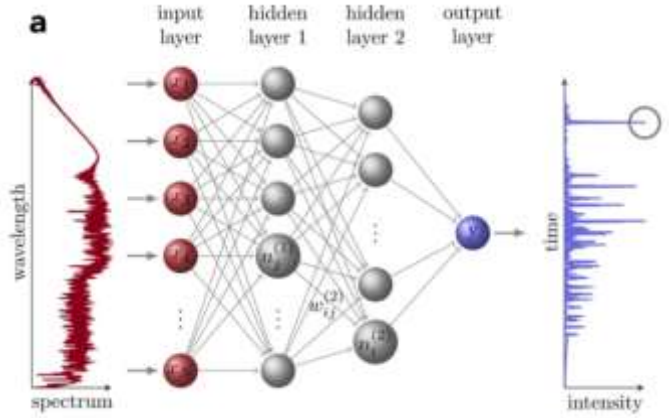



SCIENTIFIC REPORTS | (2020) 10:9596 | https://doi.org/10.1038/s41598-020-66308-y

## Machine learning analysis of rogue solitons in supercontinuum generation

Lauri Salmela[1✉], Coraline Lapre[2], John M. Dudley[2] & Goëry Genty[1]



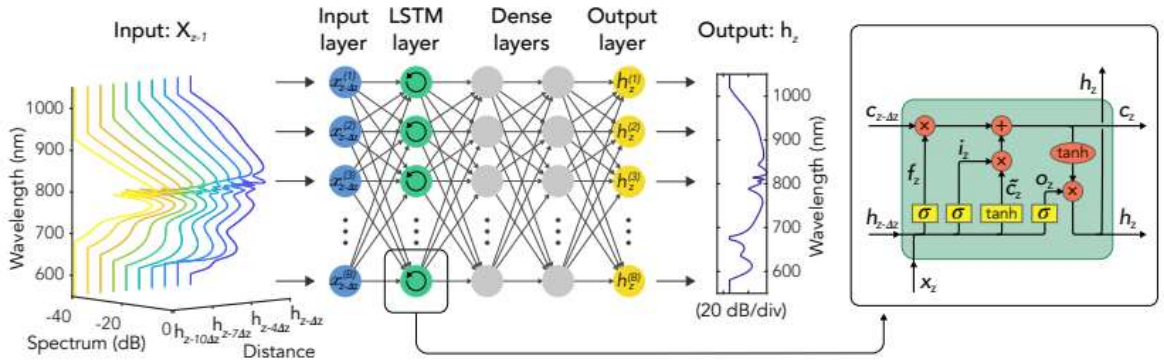

**Also:** Optics Express **30**, 15060-15072 (2022), Opt. Commun. **541**, 129570 (2023)

# Surrogate modelling of the GNLSE

Numerical integration of the generalized nonlinear Schrödinger equation (GNLSE) can be time-consuming

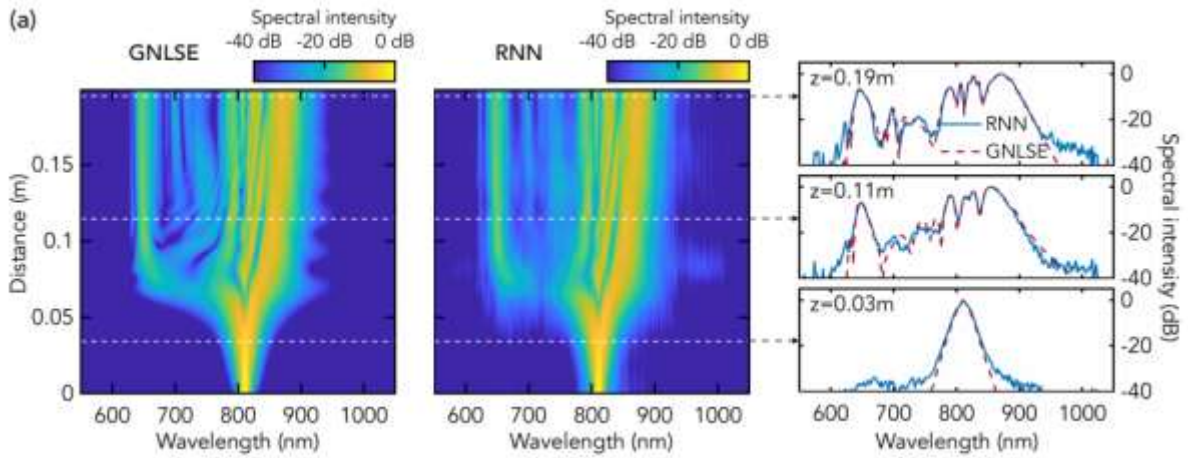
$$\frac{\partial A}{\partial z}+\beta_1\frac{\partial A}{\partial t}+\frac{i\beta_2}{2!}\frac{\partial^2 A}{\partial t^2}-\frac{\beta_3}{3!}\frac{\partial^3 A}{\partial t^3}+\ldots=i\gamma\left(1+\frac{i}{\omega_o}\frac{\partial}{\partial t}\right)\left(A(z,t)\int_{-\infty}^{\infty}R(t')\left|A(z,t-t')\right|^2dt'\right)$$

We have shown that a recurrent neural network (RNN) can replace direct numerical integration




Training uses spectral or temporal profiles from the 10 previous distance steps to predict the next profile in the evolution.

Comparing GNLSE & RNN Spectral Evolution




NATURE MACHINE INTELLIGENCE | VOL 3 | APRIL 2021 | 344–354

Good prediction within the training set of pulse & fibre parameters. Poor generalization outside.

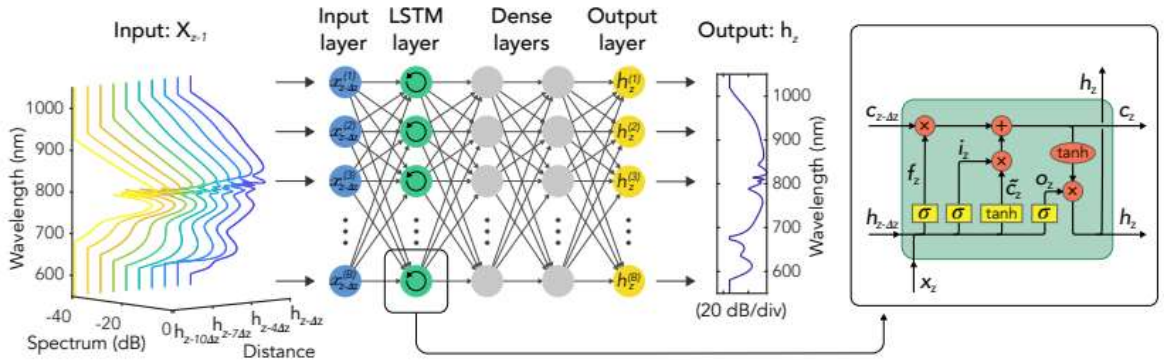
gitlab.com/salmelal/rnnnonlinear et doi.org/10.5281/zenodo.4304771 also Optics Letters 47 802-805 (2022)

# Surrogate modelling of the GNLSE

Numerical integration of the generalized nonlinear Schrödinger equation (GNLSE) can be time-consuming

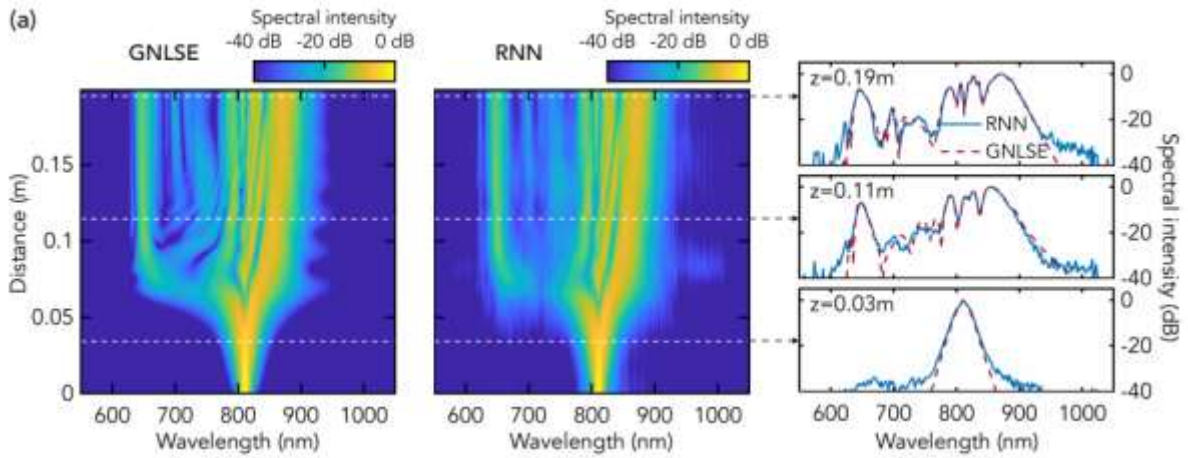
$$\frac{\partial A}{\partial z}+\beta_1\frac{\partial A}{\partial t}+\frac{i\beta_2}{2!}\frac{\partial^2 A}{\partial t^2}-\frac{\beta_3}{3!}\frac{\partial^3 A}{\partial t^3}+\ldots=i\gamma\left(1+\frac{i}{\omega_o}\frac{\partial}{\partial t}\right)\left(A(z,t)\int_{-\infty}^{\infty}R(t')\left|A(z,t-t')\right|^2dt'\right)$$

We have shown that a recurrent neural network (RNN) can replace direct numerical integration




Training uses spectral or temporal profiles from the 10 previous distance steps to predict the next profile in the evolution.

Comparing GNLSE & RNN Spectral Evolution




NATURE MACHINE INTELLIGENCE | VOL 3 | APRIL 2021 | 344–354

**Surrogate models** aim to construct accurate approximations of expensive simulators

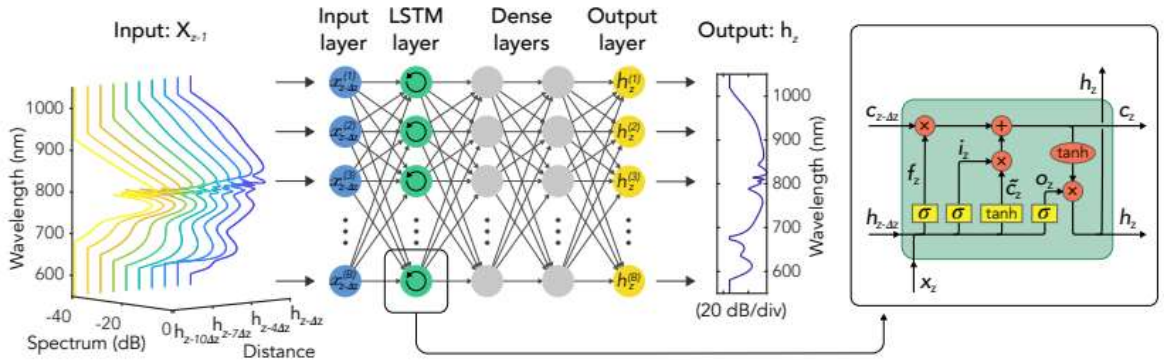
gitlab.com/salmelal/rnnnonlinear et doi.org/10.5281/zenodo.4304771 also Optics Letters 47 802-805 (2022)

# Surrogate modelling of the GNLSE

Numerical integration of the generalized nonlinear Schrödinger equation (GNLSE) can be time-consuming

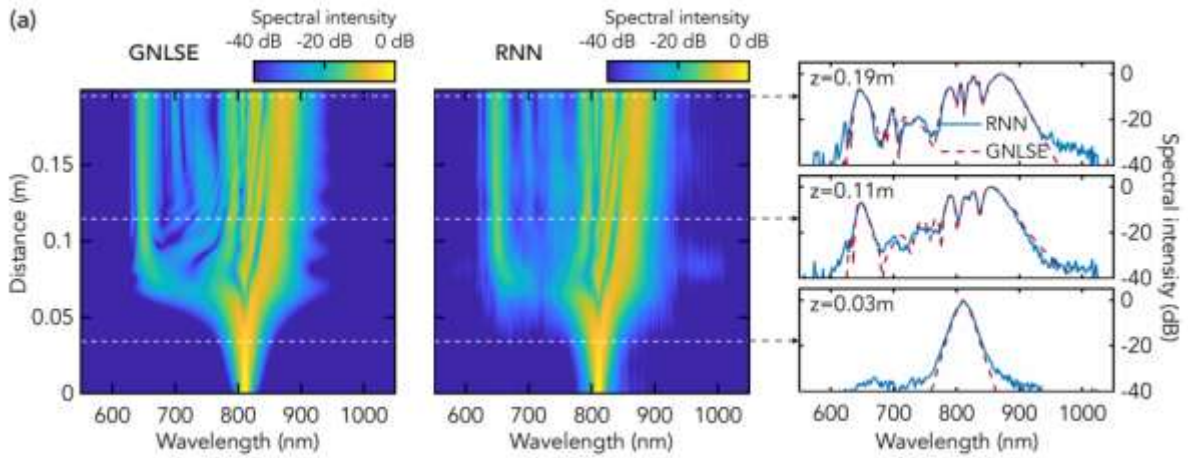
$$\frac{\partial A}{\partial z}+\beta_1\frac{\partial A}{\partial t}+\frac{i\beta_2}{2!}\frac{\partial^2 A}{\partial t^2}-\frac{\beta_3}{3!}\frac{\partial^3 A}{\partial t^3}+\ldots=i\gamma\left(1+\frac{i}{\omega_o}\frac{\partial}{\partial t}\right)\left(A(z,t)\int_{-\infty}^{\infty}R(t')\left|A(z,t-t')\right|^2dt'\right)$$

We have shown that a recurrent neural network (RNN) can replace direct numerical integration




Training uses spectral or temporal profiles from the 10 previous distance steps to predict the next profile in the evolution.

Comparing GNLSE & RNN Spectral Evolution




NATURE MACHINE INTELLIGENCE | VOL 3 | APRIL 2021 | 344–354

**Why?** For applications in optical computing or when studying supercontinuum optimization, it is sometimes necessary to perform $10^6$ simulations for one global set of parameters …

# Genetic Algorithms for fibre laser optimisation

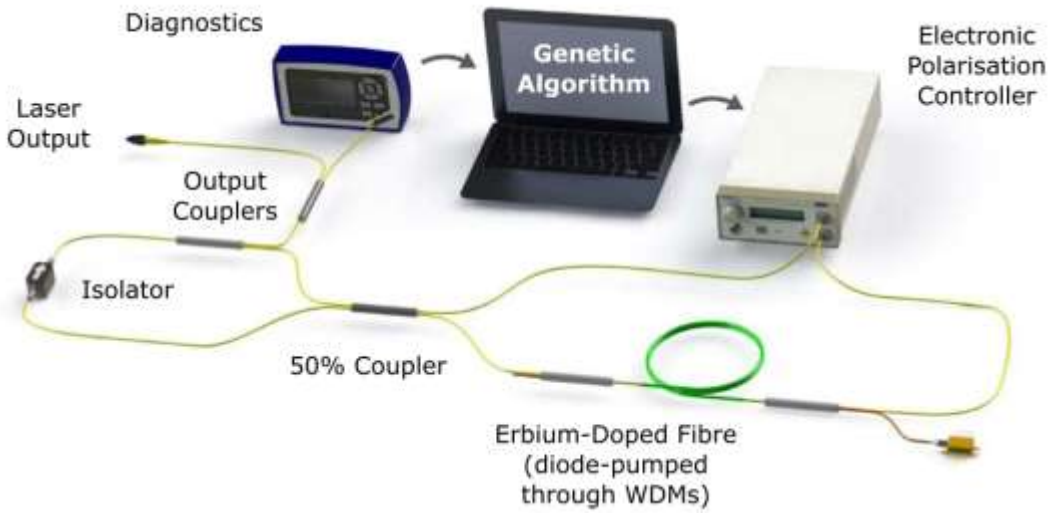
# Evolutionary or Genetic algorithms

SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

R. I. Woodward & E. J. R. Kelleher




Optimize the laser by varying orientation of 4 waveplates $\theta_1$ $\theta_2$ $\theta_3$ $\theta_4$ (orientation = gene)

When a pulse propagates in an optical fiber which is not polarization-maintaining, there can be a **nonlinear intensity-dependent** change in the polarization state.

This effect is often used for **passive mode locking** because a high-intensity pulse develops a different polarization state to low-intensity noise

A typical configuration contains a set of waveplates adjusted such that the maximum transmission (minimum loss) at the polarizer occurs for the highest possible optical intensity.

The configuration then serves as an **artificial saturable absorber.**

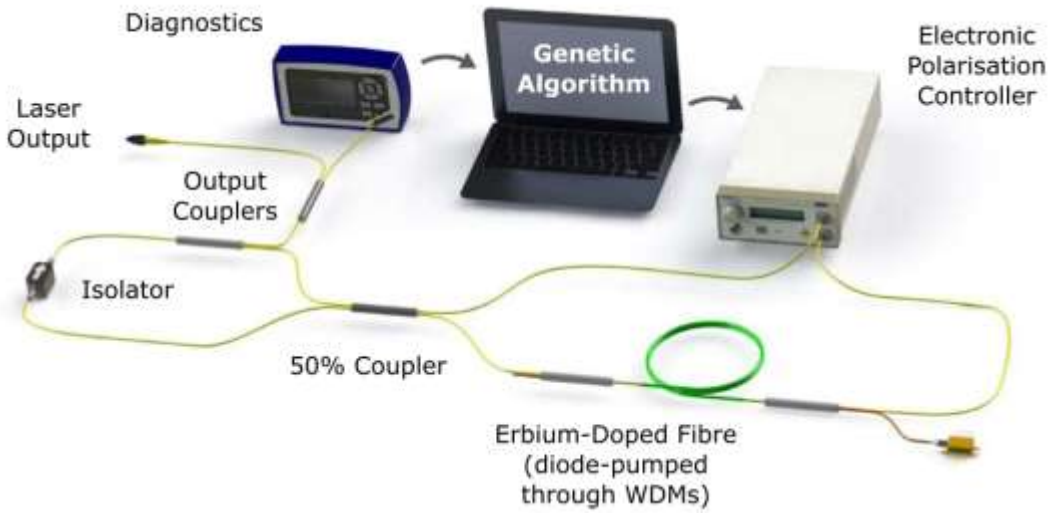
# Evolutionary or Genetic algorithms

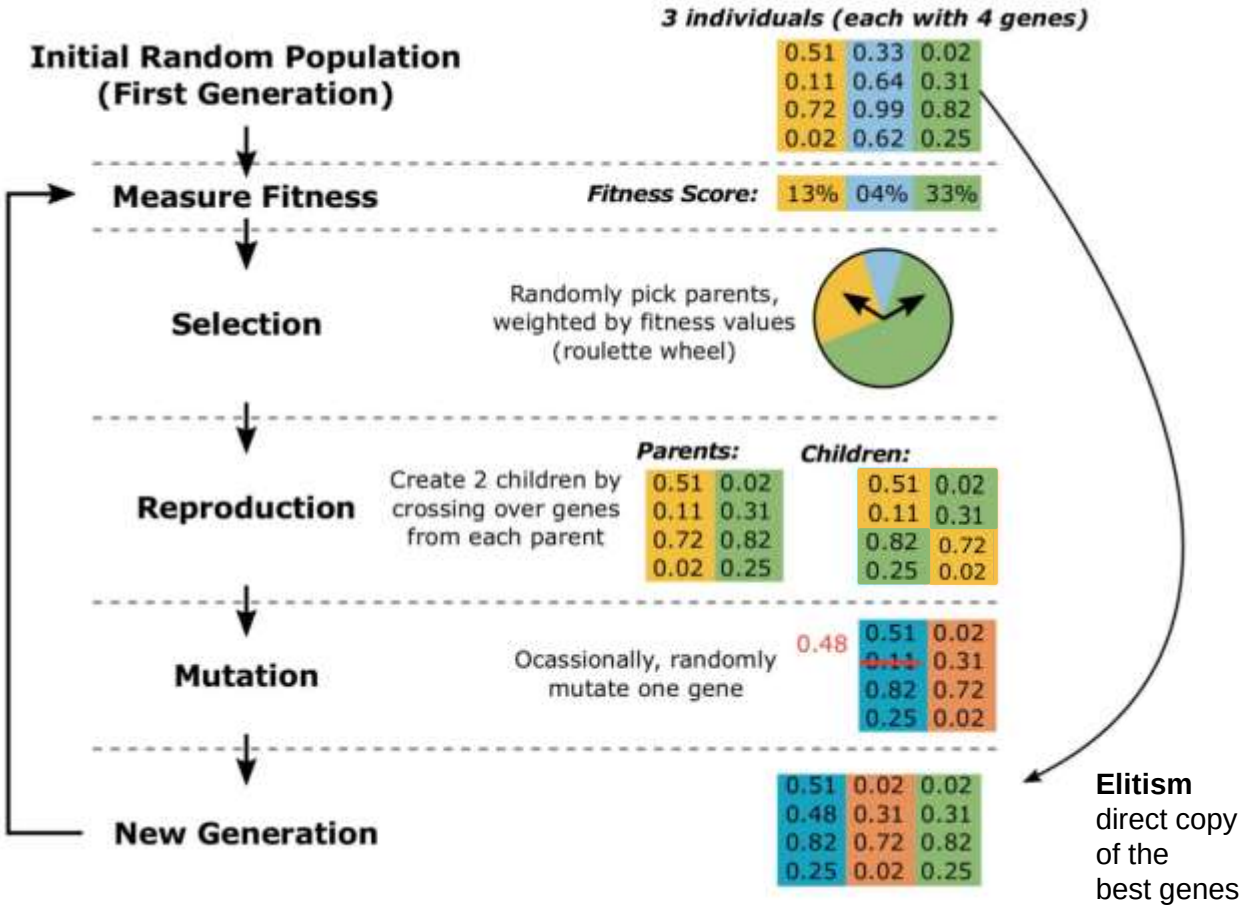
SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

R. I. Woodward & E. J. R. Kelleher




Optimize the laser by varying orientation of 4 waveplates $\theta_1$ $\theta_2$ $\theta_3$ $\theta_4$ (orientation = gene)

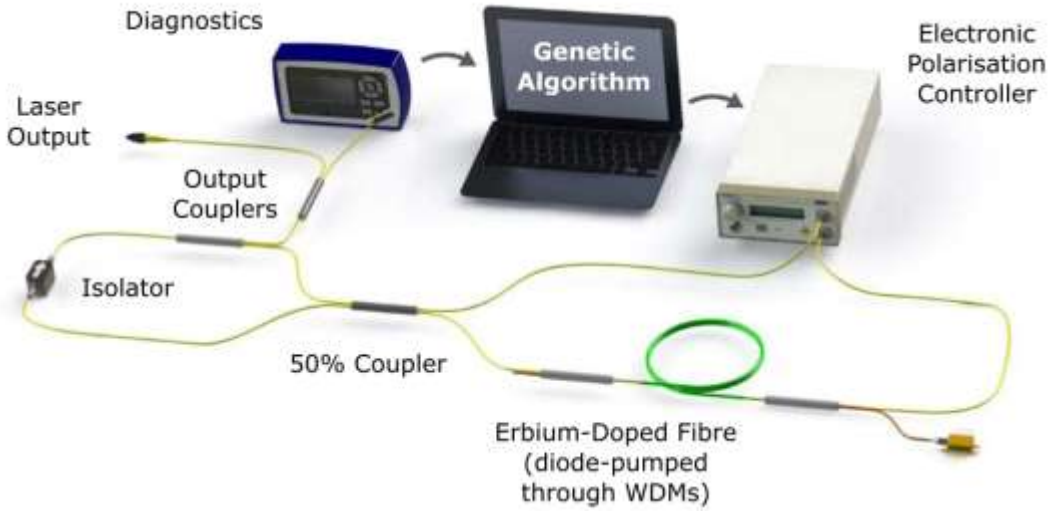
# Evolutionary or Genetic algorithms

SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

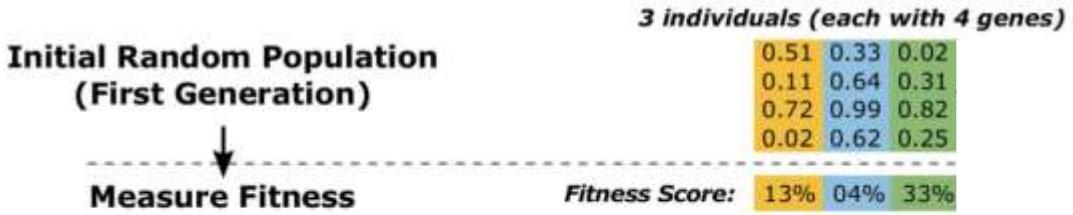
Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

R. I. Woodward & E. J. R. Kelleher




Optimize the laser by varying orientation of 4 waveplates $\theta_1$ $\theta_2$ $\theta_3$ $\theta_4$ (orientation = gene)




Each initial random set of 4 waveplate orientations forms an "individual"

| | | | |
|---|---|---|---|
| QWP | 0.51 | 0.33 | 0.02 |
| HWP | 0.11 | 0.64 | 0.31 |
| HWP | 0.72 | 0.99 | 0.82 |
| QWP | 0.02 | 0.62 | 0.25 |

These numbers $\in (0,1)$ are waveplate angles $\theta/2\pi$

Referred to as "genes" or "variables"

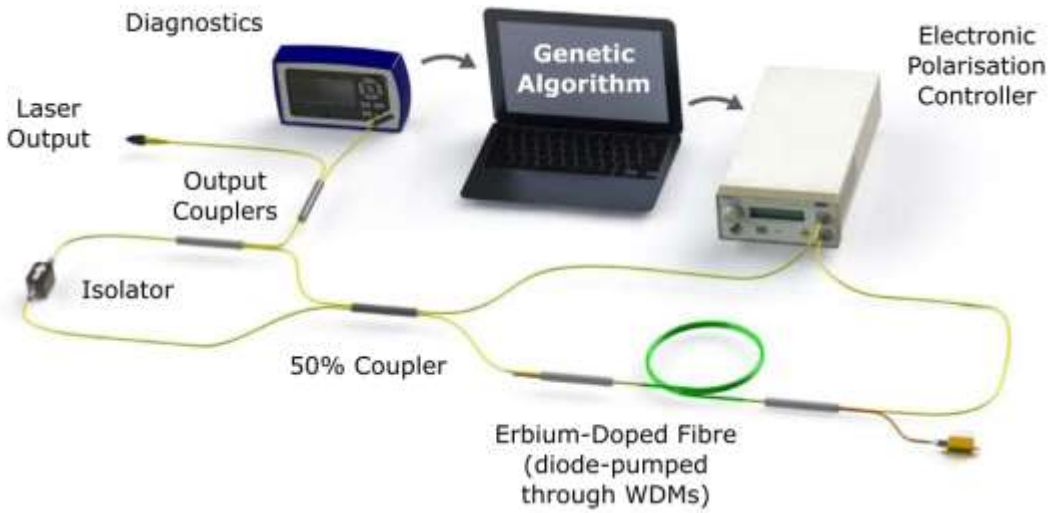
# Evolutionary or Genetic algorithms

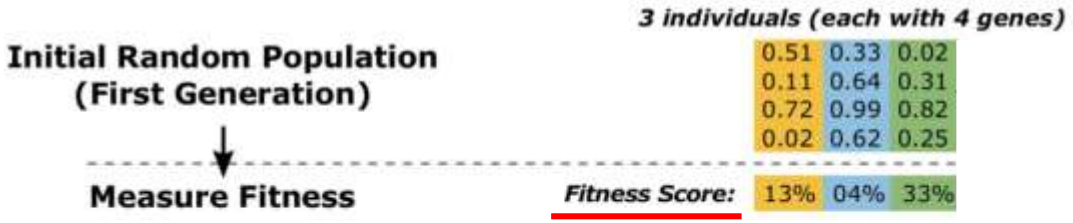
SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

R. I. Woodward & E. J. R. Kelleher




The "fitness score" or "target" or "objective function" is the experimental parameter or parameters you want to optimize

Power
Bandwidth
Autocorrelation intensity
RF spectral amplitude
….

Optimize the laser by varying orientation of 4 waveplates $\theta_1$ $\theta_2$ $\theta_3$ $\theta_4$ (orientation = gene)

A **compound** fitness score combines several of these quantities

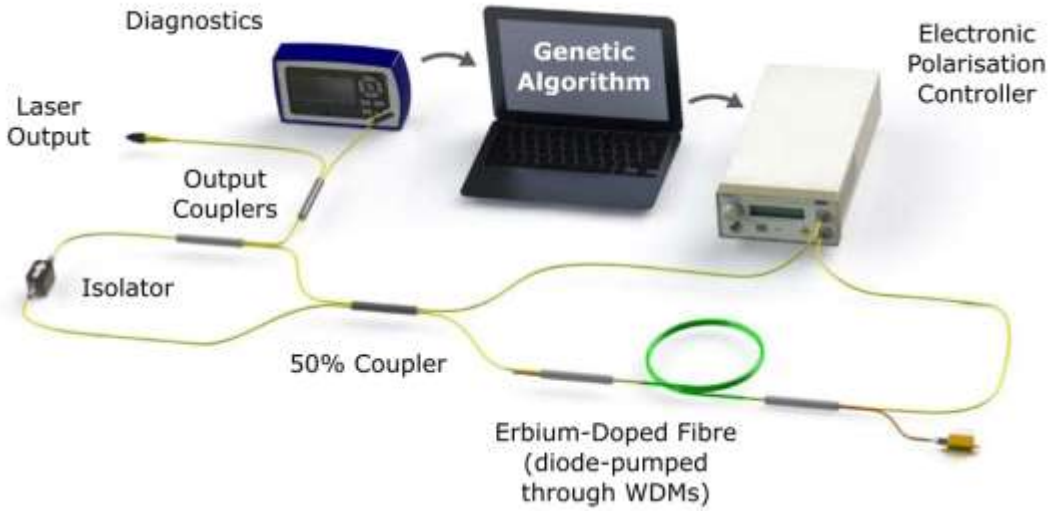
# Evolutionary or Genetic algorithms

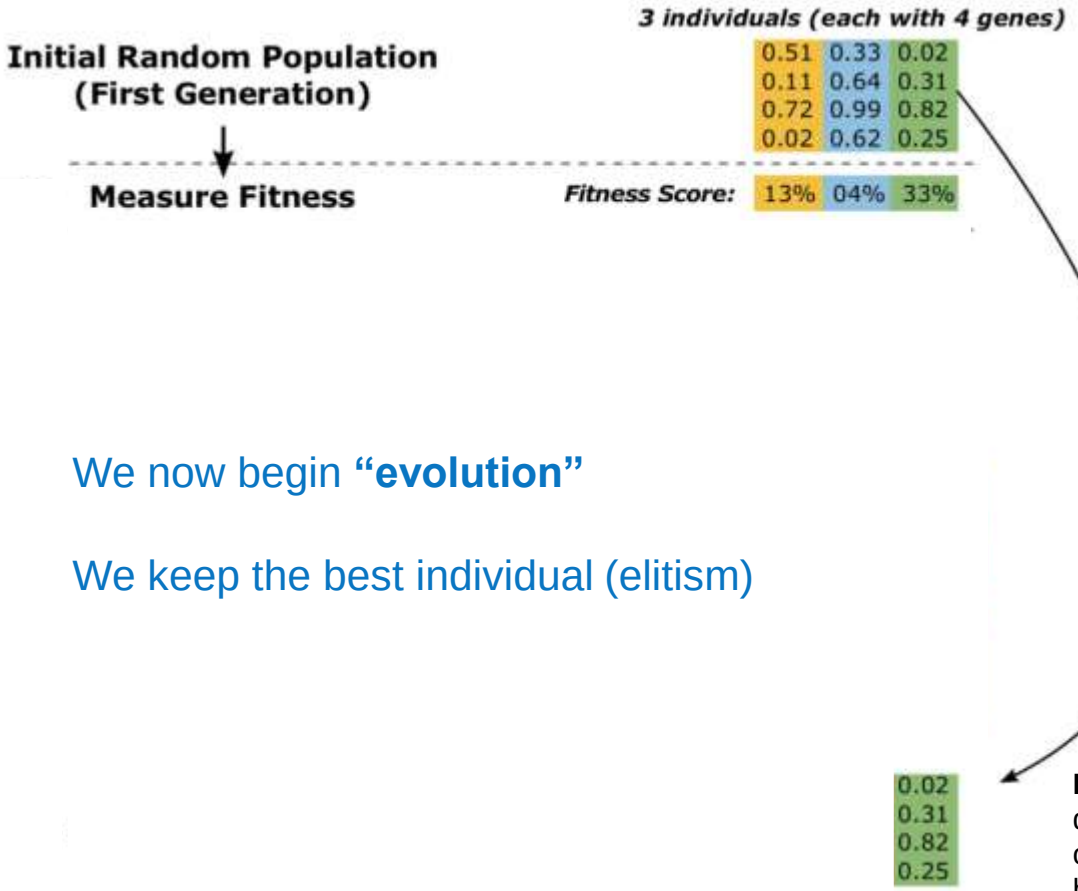
SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

R. I. Woodward & E. J. R. Kelleher




We now begin **"evolution"**

We keep the best individual (elitism)

Optimize the laser by varying orientation of 4 waveplates $\theta_1$ $\theta_2$ $\theta_3$ $\theta_4$ (orientation = gene)

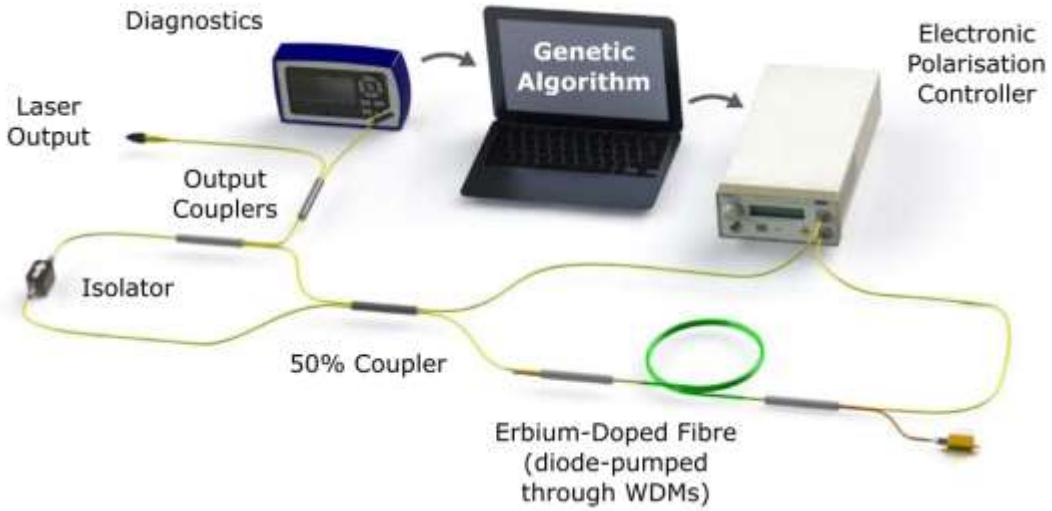
# Evolutionary or Genetic algorithms

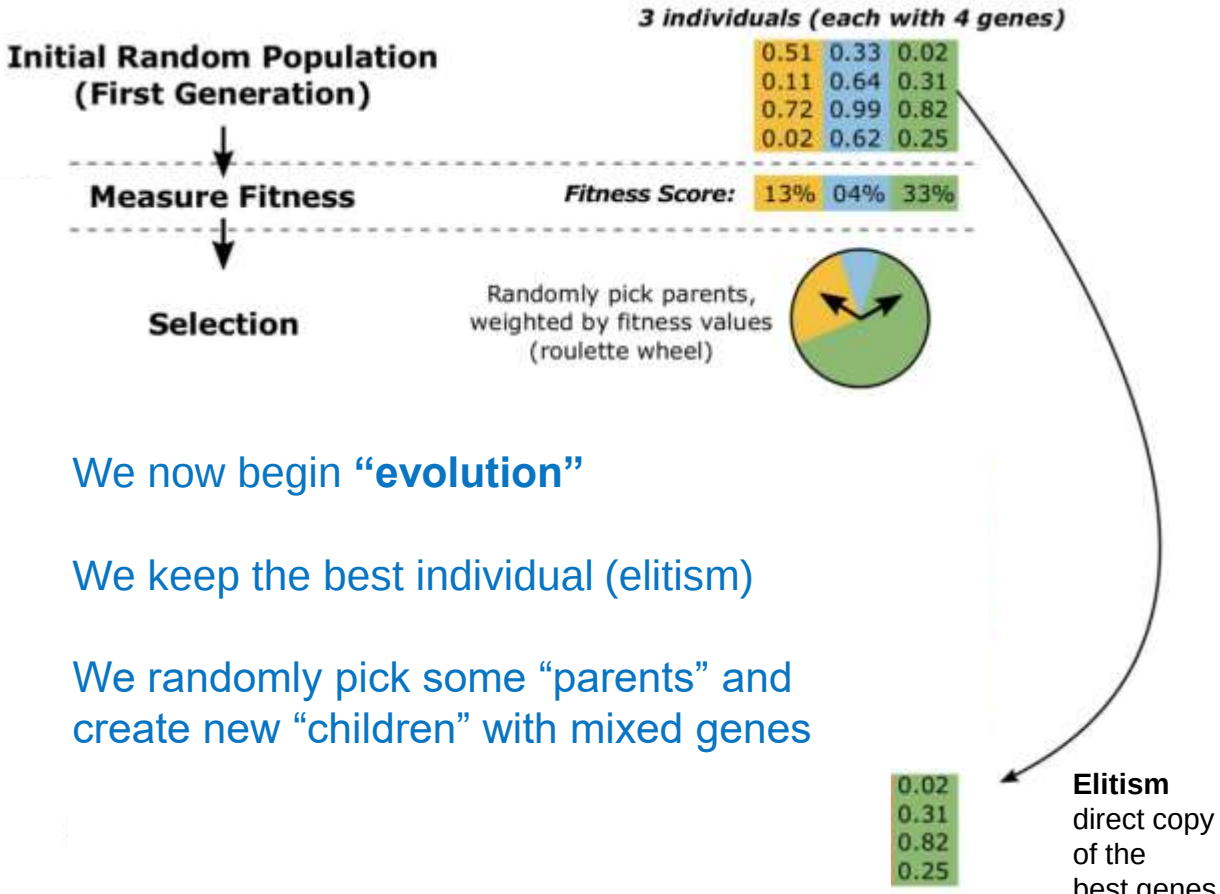
SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

R. I. Woodward & E. J. R. Kelleher




We now begin **"evolution"**

We keep the best individual (elitism)

We randomly pick some "parents" and create new "children" with mixed genes

Optimize the laser by varying orientation of 4 waveplates $\theta_1$ $\theta_2$ $\theta_3$ $\theta_4$ (orientation = gene)

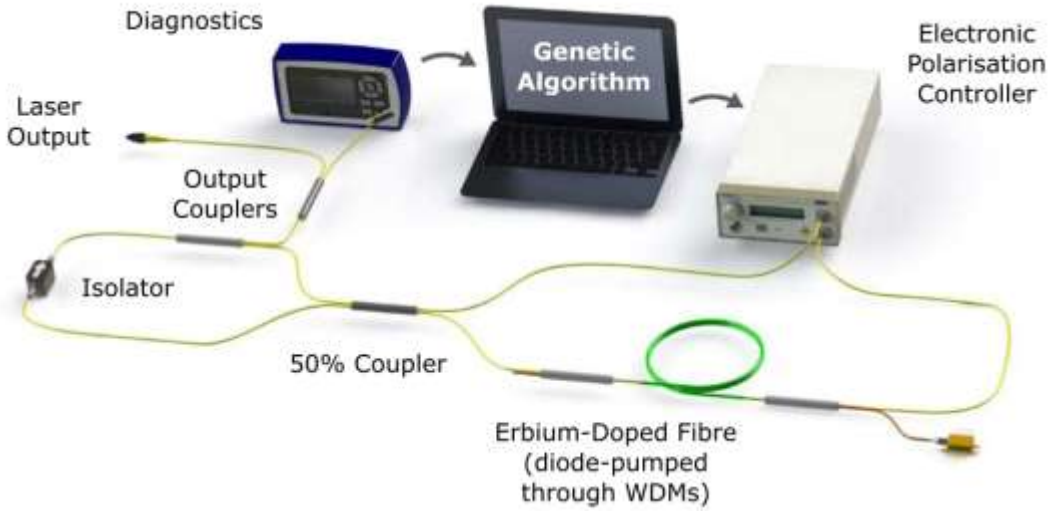
# Evolutionary or Genetic algorithms

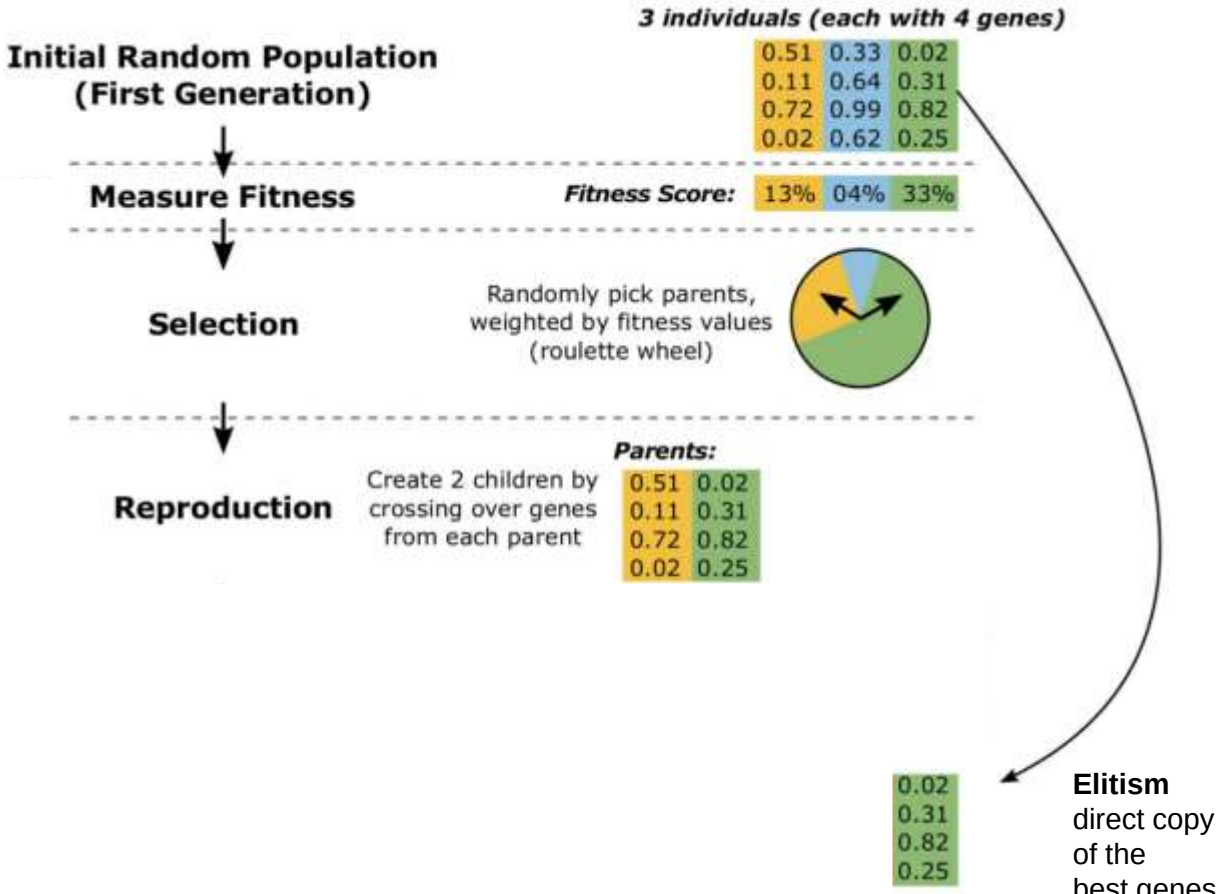
SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

R. I. Woodward & E. J. R. Kelleher




Optimize the laser by varying orientation of 4 waveplates $\theta_1$ $\theta_2$ $\theta_3$ $\theta_4$ (orientation = gene)

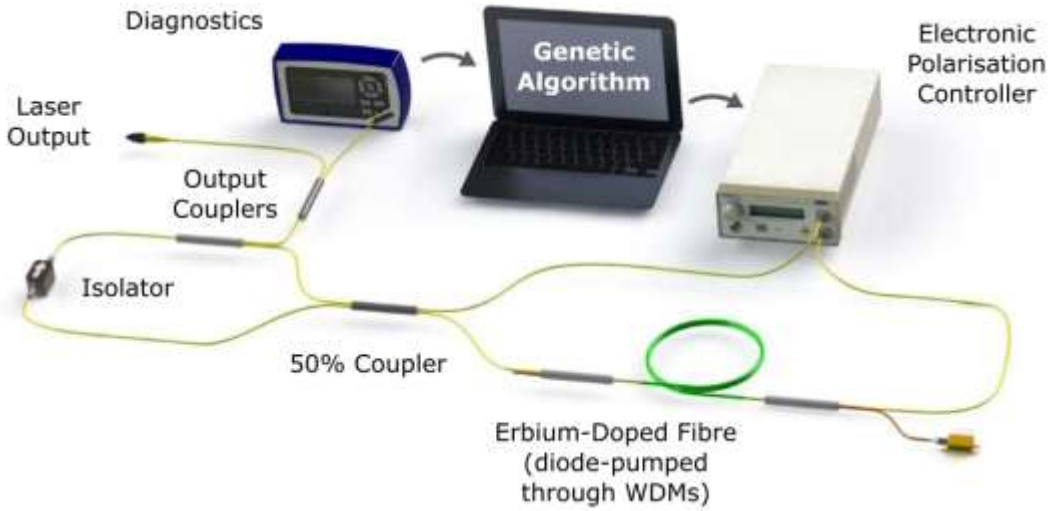
# Evolutionary or Genetic algorithms

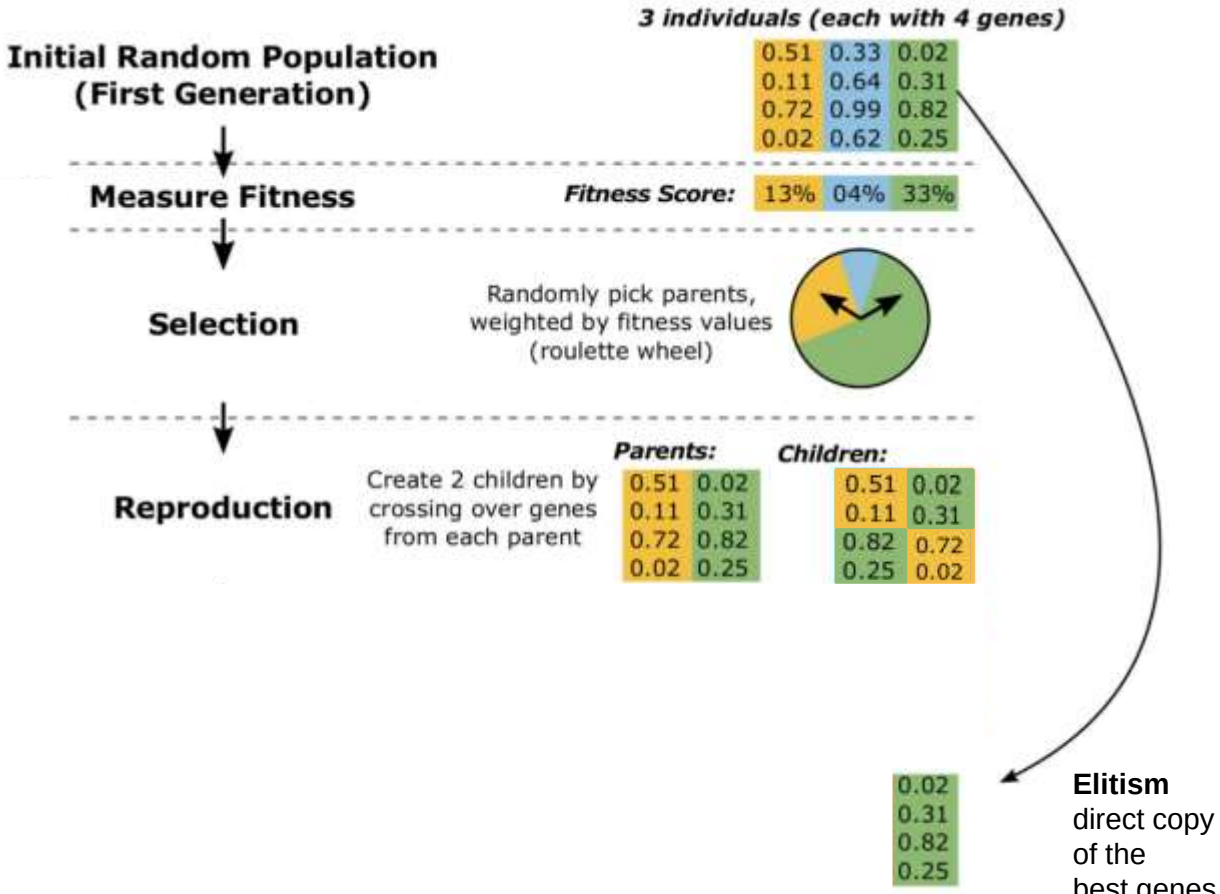
SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

R. I. Woodward & E. J. R. Kelleher




Optimize the laser by varying orientation of 4 waveplates $\theta_1$ $\theta_2$ $\theta_3$ $\theta_4$ (orientation = gene)

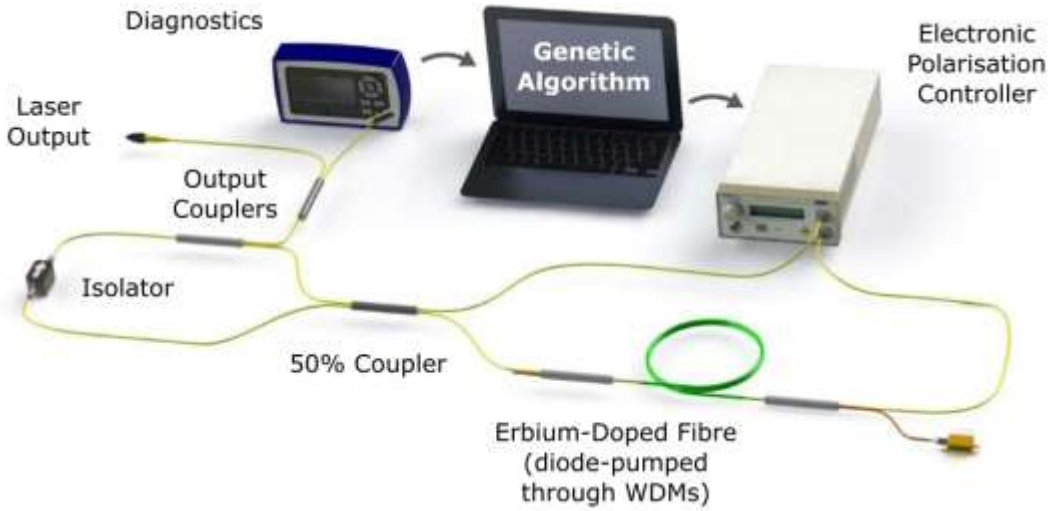
# Evolutionary or Genetic algorithms

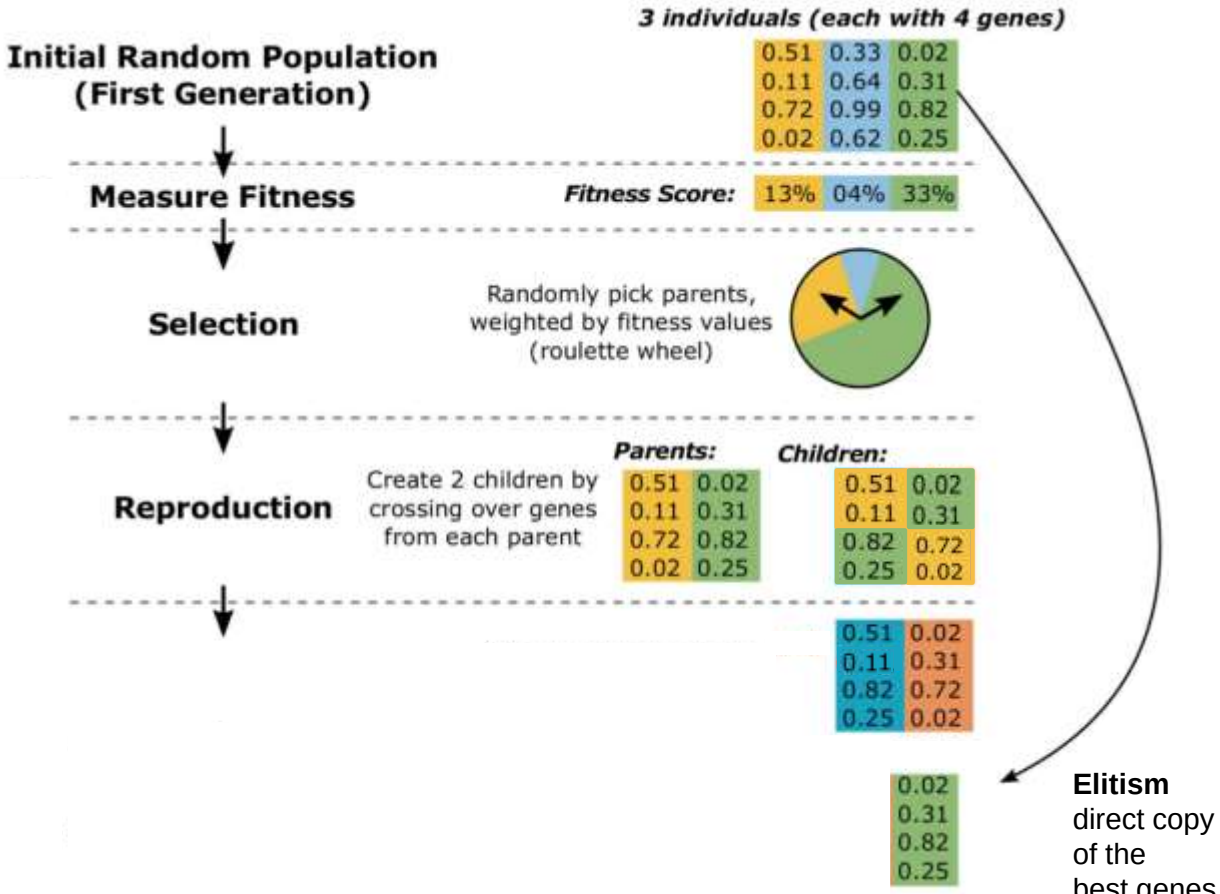
SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

R. I. Woodward & E. J. R. Kelleher




Optimize the laser by varying orientation of 4 waveplates (orientation = gene)

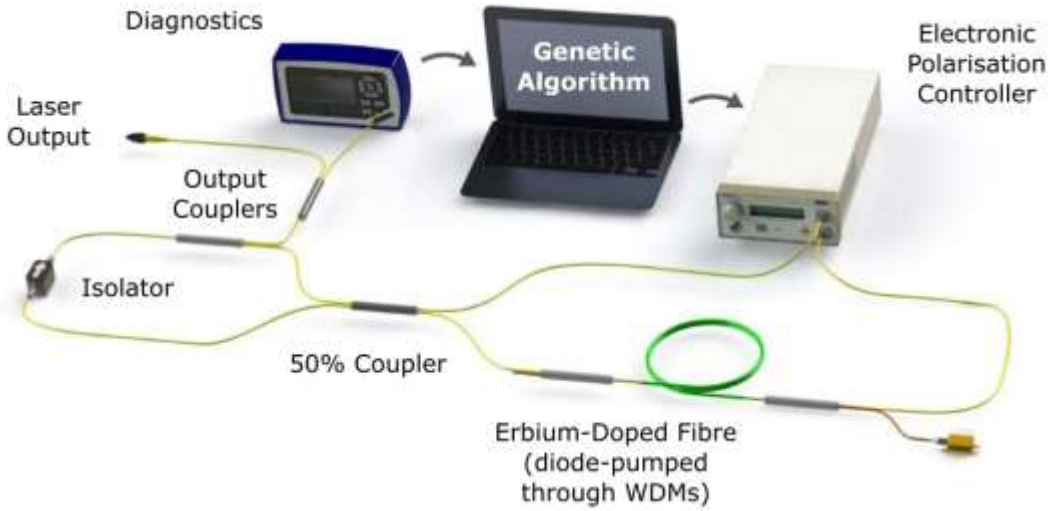
# Evolutionary or Genetic algorithms

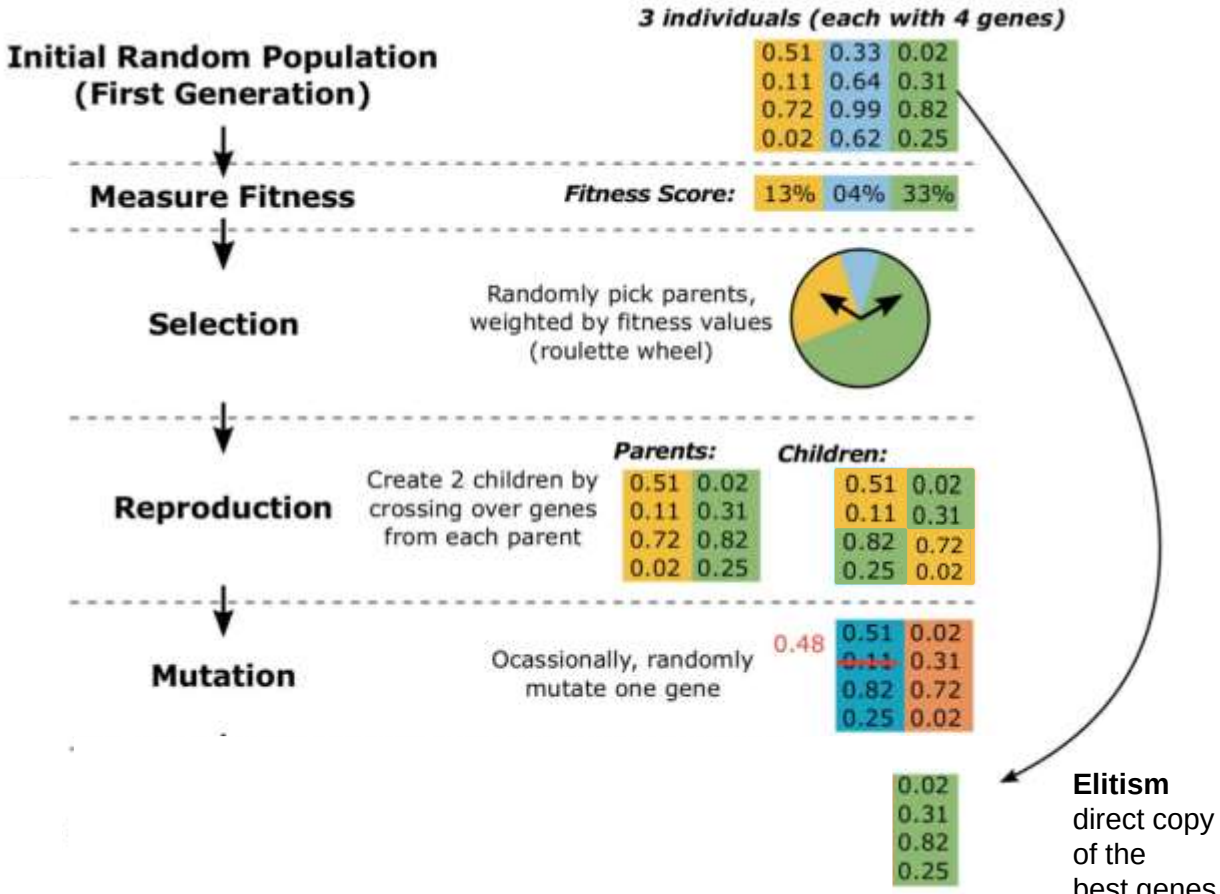
SCIENTIFIC REPORTS | 6:37616 | DOI: 10.1038/srep37616

Published: 21 November 2016

**Towards 'smart lasers': self-optimisation of an ultrafast pulse source using a genetic algorithm**

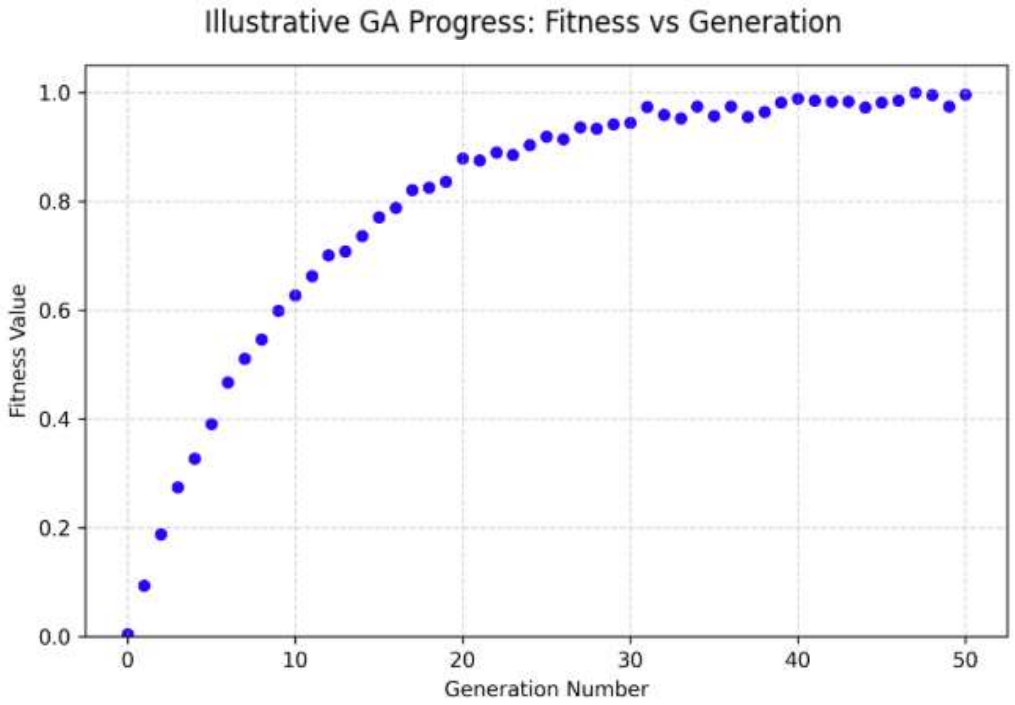
R. I. Woodward & E. J. R. Kelleher

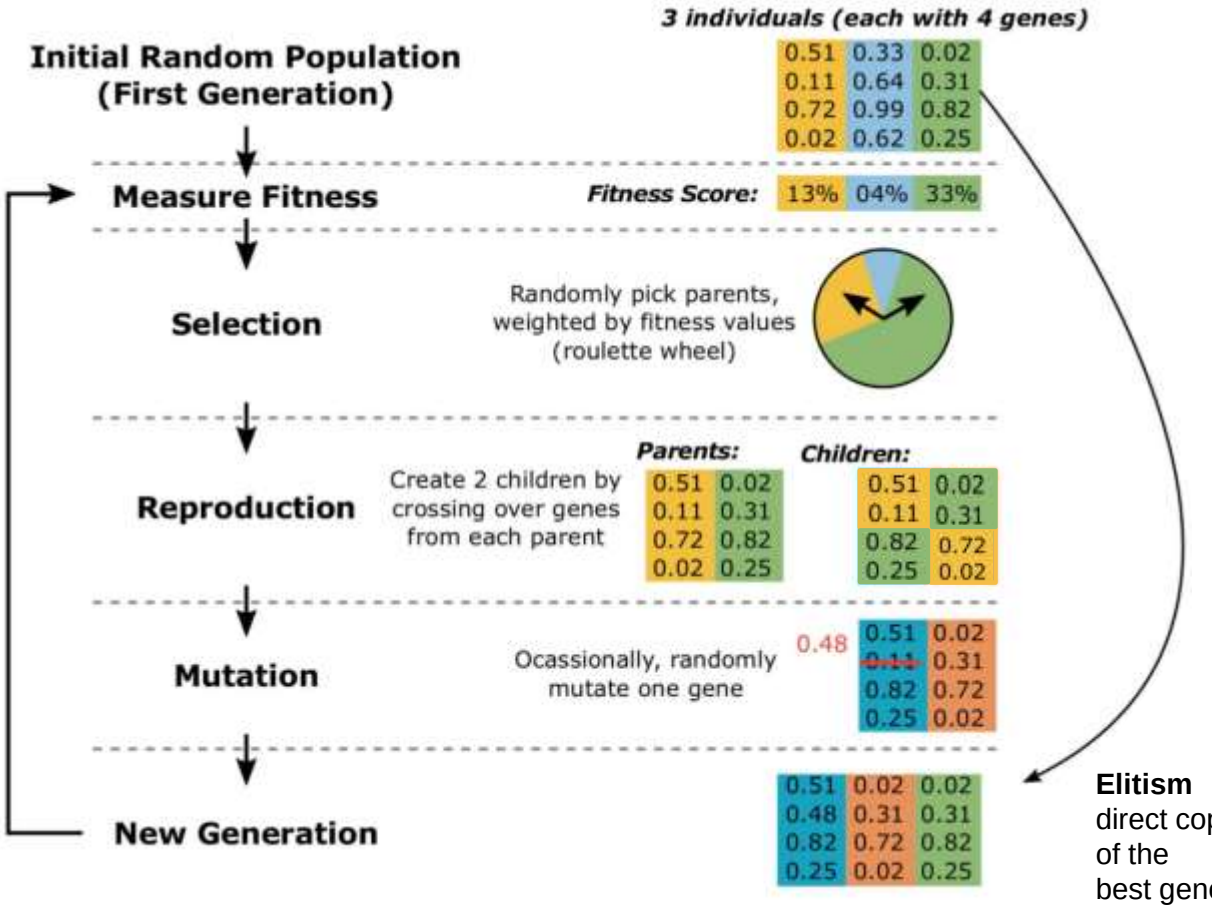



Optimize the laser by varying orientation of 4 waveplates $\theta_1$ $\theta_2$ $\theta_3$ $\theta_4$ (orientation = gene)

# Evolutionary or Genetic algorithms



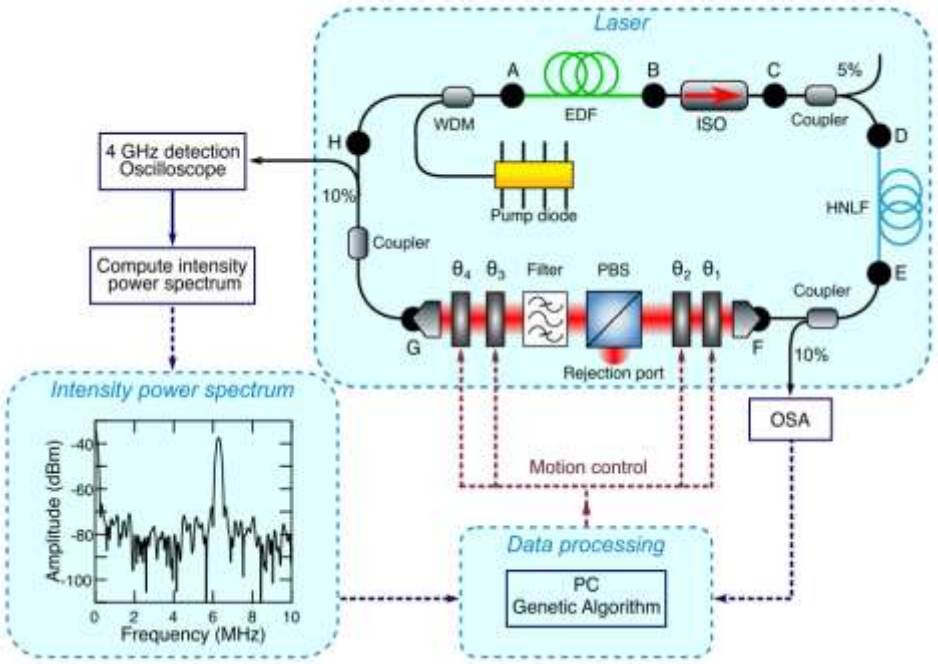



# Genetic algorithm optimization of a noise-like pulse laser

Scientific Reports | (2023) 13:1865

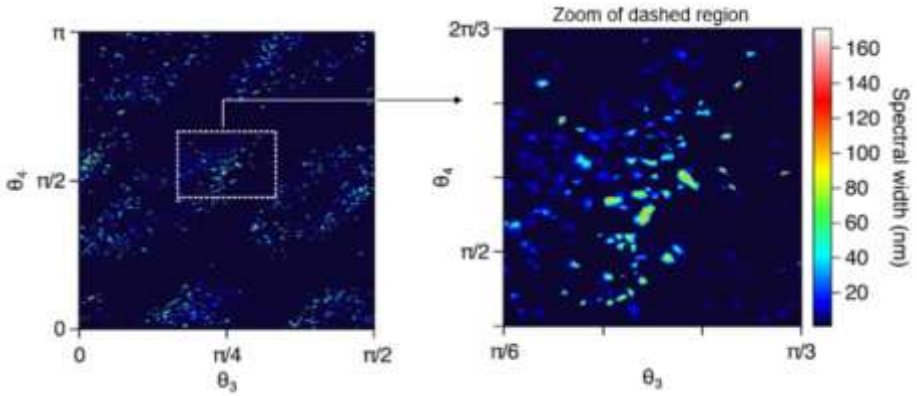
## Genetic algorithm optimization of broadband operation in a noise-like pulse fiber laser

Coraline Lapre[1], Fanchao Meng[2], Mathilde Hary[1,3], Christophe Finot[4], Goëry Genty[3] & John M. Dudley[1]

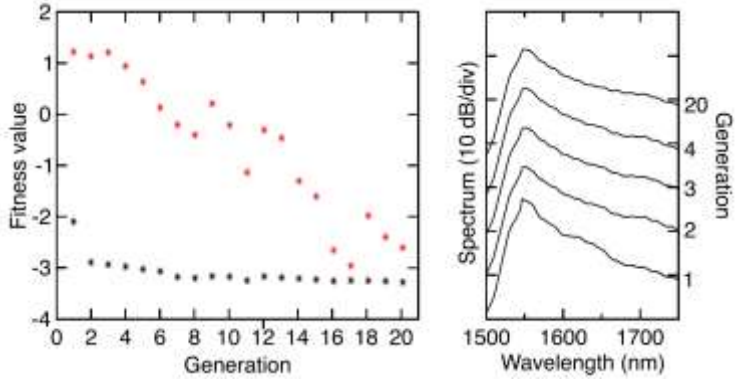



Noise like lasers show highly-complex dynamics with 100-1000 sub-ps pulses evolving randomly underneath a ps-ns envelope

The broadest bandwidths occupy a very small subspace of intracavity polarization settings, time-consuming to locate manually




A GA rapidly finds the broadband states



# Tailored supercontinuum – time domain initial conditions

Tailoring supercontinuum generation by active control of initial conditions is a challenging goal. One early paper used programmable cascaded Mach Zehnder interferometers to generate multiple pulse patterns optimized using a genetic algorithm to favour enhancement at particular wavelengths.

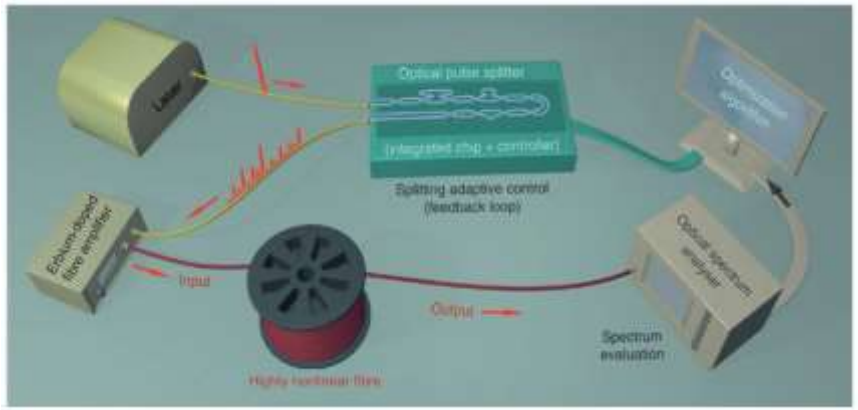
NATURE COMMUNICATIONS | (2018)9:4884

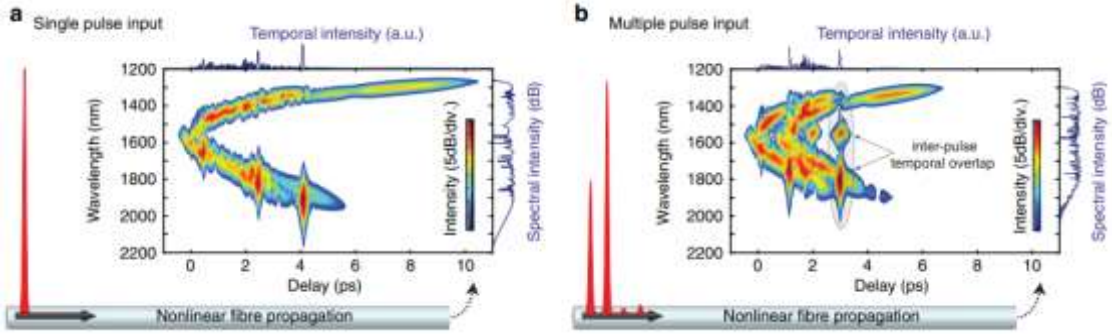
ARTICLE

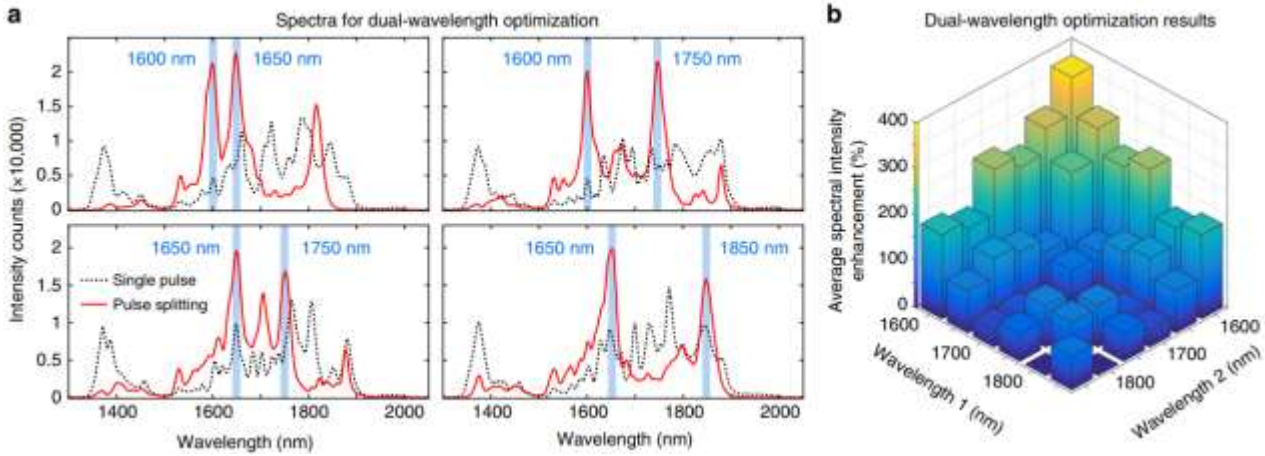
DOI: 10.1038/s41467-018-07143-w OPEN

## Customizing supercontinuum generation via on-chip adaptive temporal pulse-splitting

Benjamin Wetzel[1,2], Michael Kues[1,3], Piotr Roztocki[1], Christian Reimer[1,4], Pierre-Luc Godin[1], Maxwell Rowley[2], Brent E. Little[5], Sai T. Chu[6], Evgeny A. Viktorov[7], David J. Moss[8], Alessia Pasquazi[2], Marco Peccianti[2] & Roberto Morandotti[1,7,9]

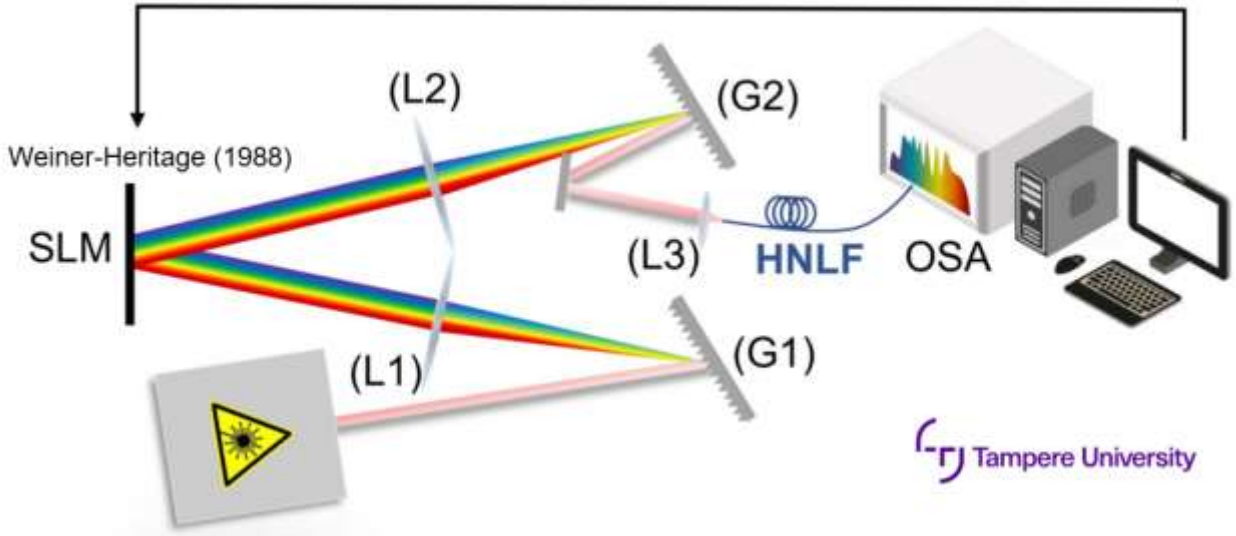


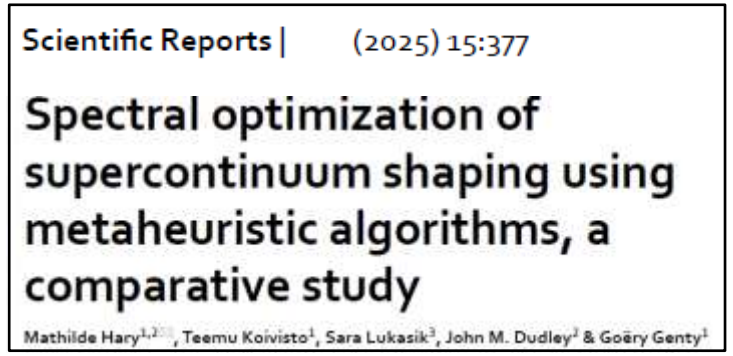



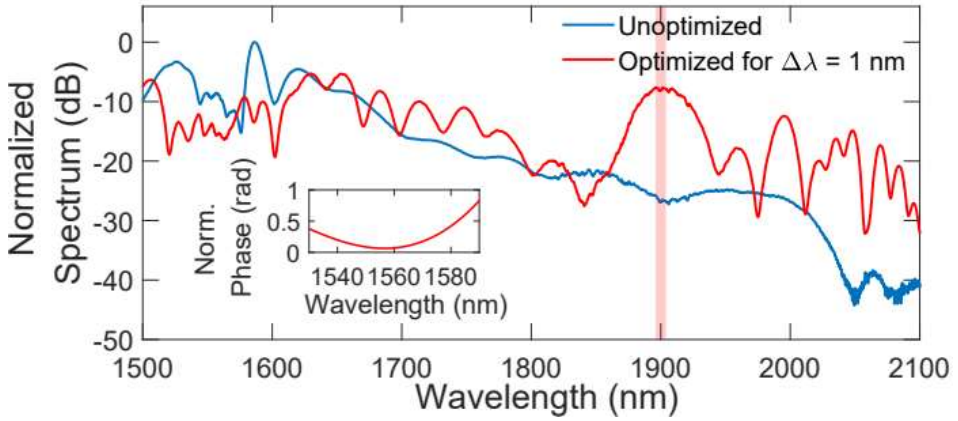
# Tailored supercontinuum – frequency domain initial conditions


$$\Phi_{SLM}(\omega) = c_2(\omega - c_5)^2 + c_3(\omega - c_5)^3 + c_4(\omega - c_5)^4$$

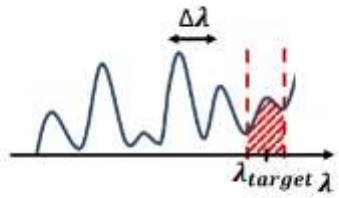



Enhance intensity at a desired wavelength




SINGLE WAVELENGTH FITNESS FUNCTION



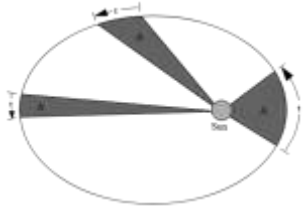

$$S = \int_{\Delta\lambda} I(\lambda) d\lambda$$

# Can AI discover models from data ?

# Data driven model discovery

Data driven discovery describes  finding new scientific laws, patterns, or phenomena by analyzing data first, often before a full theory exists

## 1609 Kepler's laws

$$T^2 \propto r^3$$

## 1885 Balmer series of H-spectrum

**V. *Notiz über die Spectrallinien des Wasserstoffs***
***von J. J. Balmer.***

Basel, 30. Januar 1885.

$$\frac{1}{\lambda} = R\left(\frac{1}{2^2} - \frac{1}{n^2}\right) \quad , \quad n = 3,4,5,6,\ldots$$

$$R = 10973731.57\,\mathrm{m}^{-1}$$

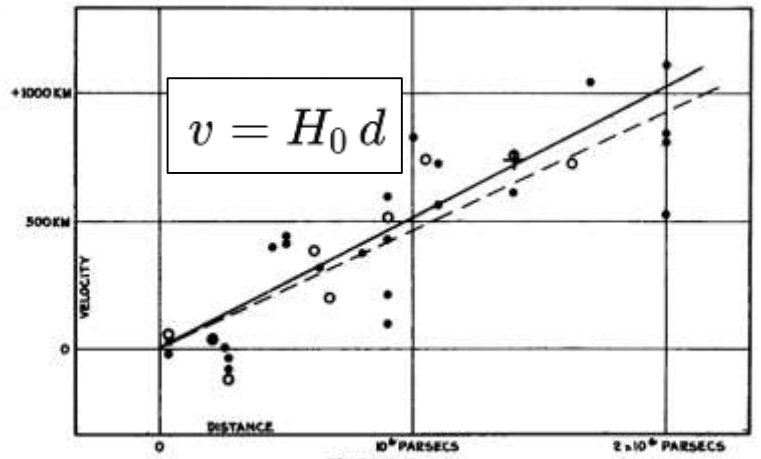
## 1929 Hubble's law

*A RELATION BETWEEN DISTANCE AND RADIAL VELOCITY AMONG EXTRA-GALACTIC NEBULAE*

By Edwin Hubble

Mount Wilson Observatory, Carnegie Institution of Washington

Communicated January 17, 1929




$$v = H_0\, d$$

FIGURE 1

Velocity-Distance Relation among Extra-Galactic Nebulae.

# Data driven model discovery

In 2016, a new algorithmic method **SINDy** was introduced, aiming to "reverse engineer" nonlinear dynamical time series or space series data so as to determine the governing equations.

PNAS 3932–3937 | PNAS | April 12, 2016 | vol. 113 | no. 15

**Discovering governing equations from data by sparse identification of nonlinear dynamical systems**

Steven L. Brunton[a,1], Joshua L. Proctor[b], and J. Nathan Kutz[c]

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \frac{d}{d\zeta}\mathbf{x}(\zeta) = \mathbf{f}[\mathbf{x}(\zeta)]$$

State vector (data is known)

Function defining equations of motion (unknown)

Time or propagation coordinate

We guess candidate terms in the equations of motion and use regression to determine coefficients of each term to best fit the data (i.e. we minimise the error between the guess and the data).

# Data driven model discovery

In 2016, a new algorithmic method **SINDy** was introduced, aiming to “reverse engineer” nonlinear dynamical time series or space series data so as to determine the governing equations.

PNAS 3932–3937 | PNAS | April 12, 2016 | vol. 113 | no. 15

**Discovering governing equations from data by sparse identification of nonlinear dynamical systems**

Steven L. Brunton[a,1], Joshua L. Proctor[b], and J. Nathan Kutz[c]

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

State vector (data is known)

Guessed candidate functions

$$\dot{X} = \Theta(X)\ C$$

Coefficients


We guess candidate terms in the equations of motion and use regression to determine coefficients of each term to best fit the data (i.e. we minimise the error between the guess and the data).

# Example: SINDy applied to the “Lorenz 63” system

A toy model for convection in the atmosphere

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x),$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$

**x:** a convective flow variable
**y:** related to horizontal distribution of temperature
**z:** related to vertical distribution of temperature

**σ:** ratio between viscosity and thermal conductivity
ρ: temperature difference (Rayleigh)
β: ratio between width and height (geometric)

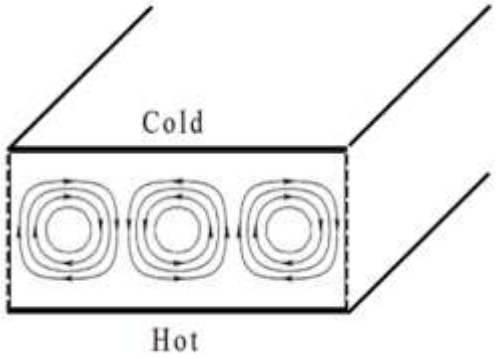
JOURNAL OF THE ATMOSPHERIC SCIENCES

**Deterministic Nonperiodic Flow**[1]

EDWARD N. LORENZ

*Massachusetts Institute of Technology*

(Manuscript received 18 November 1962, in revised form 7 January 1963)



# Example: SINDy applied to the “Lorenz 63” system

A toy model for convection in the atmosphere

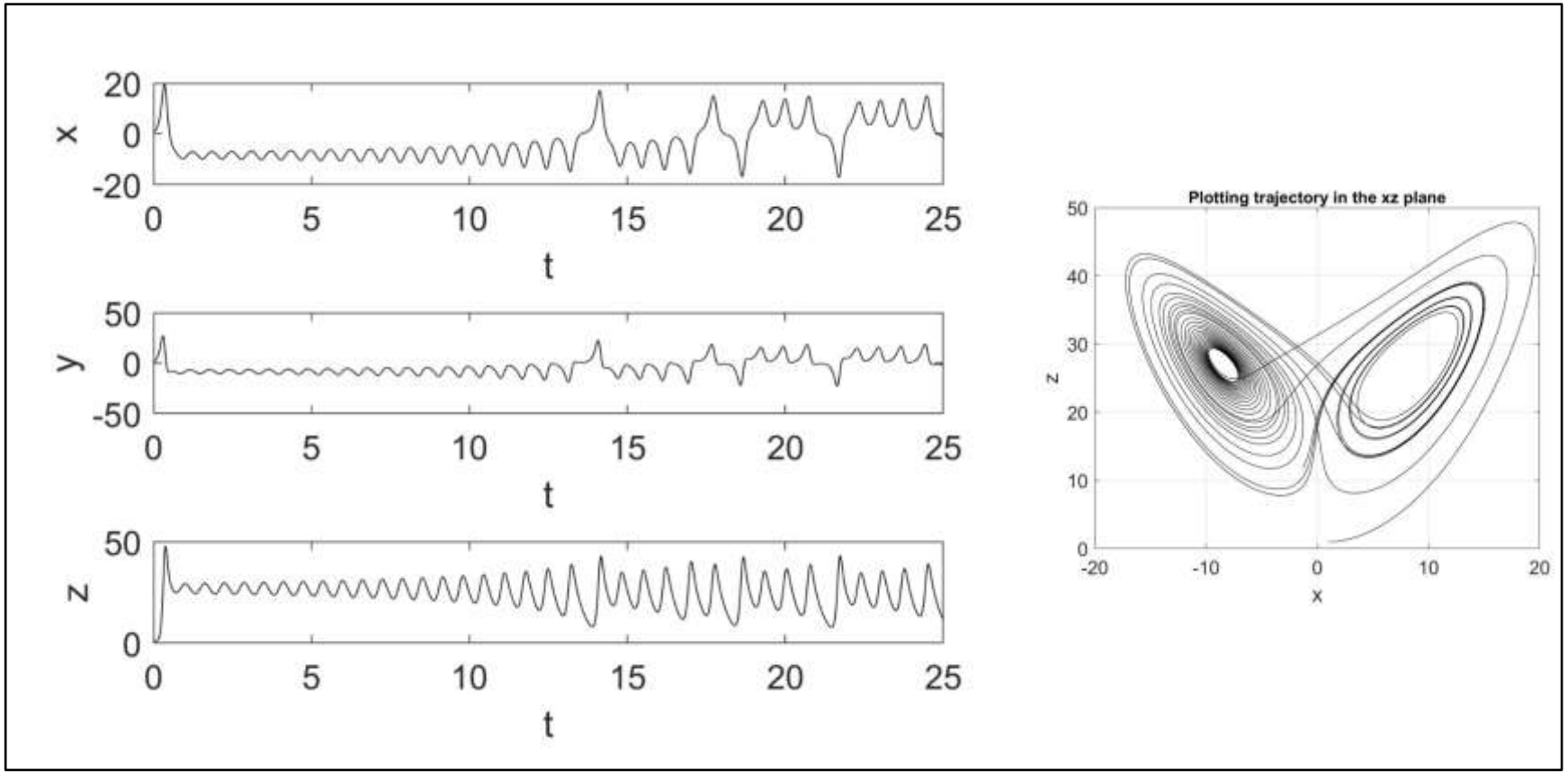
$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x),$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$

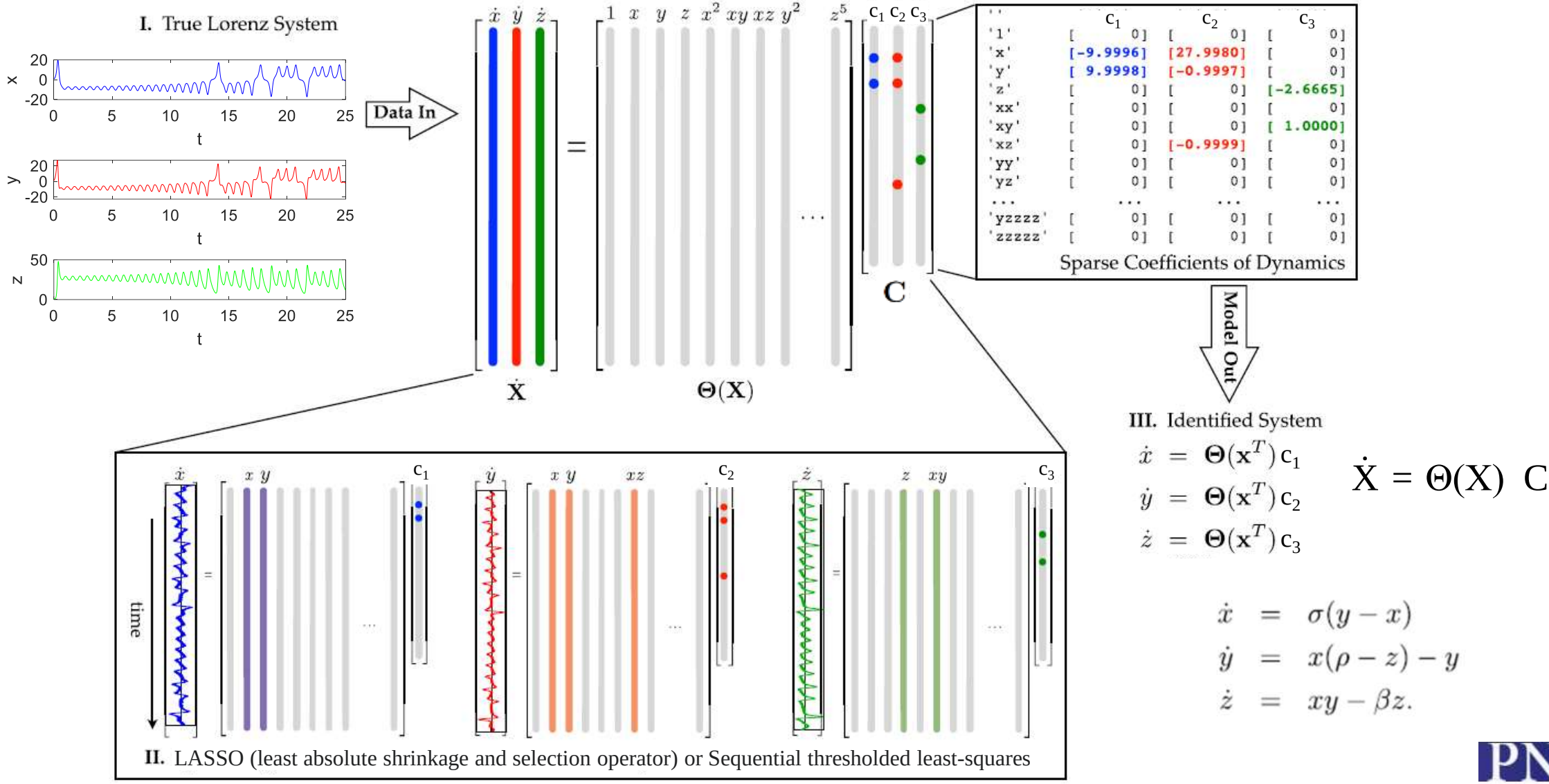
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$

```
% parameters      % Initial conditions

sigma = 10;       x0 = 1;
beta = 8/3;       y0 = 1;
rho = 28;         z0 = 1;
```



# Example: SINDy applied to the Lorenz 63 system




$$\dot{X} = \Theta(X)\ C$$

$$\begin{aligned}\dot{x} &= \sigma(y-x)\\ \dot{y} &= x(\rho-z)-y\\ \dot{z} &= xy-\beta z.\end{aligned}$$

Steven L. Brunton[a,1], Joshua L. Proctor[b], and J. Nathan Kutz[c]

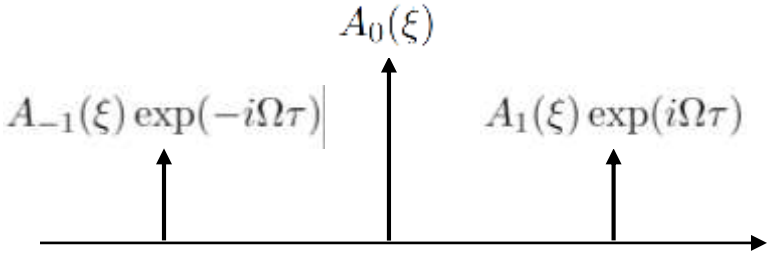
PNAS

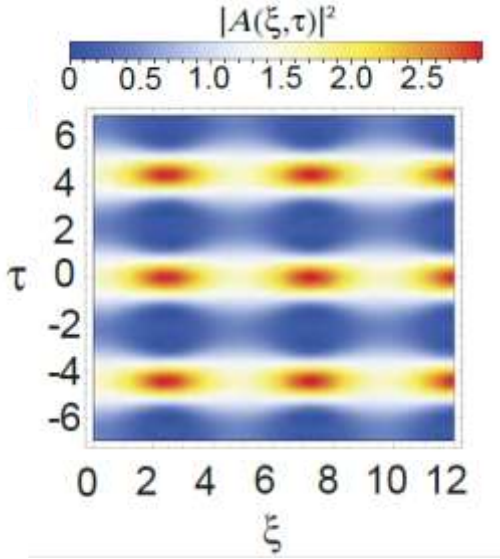
# SINDy applied to four wave mixing in optics

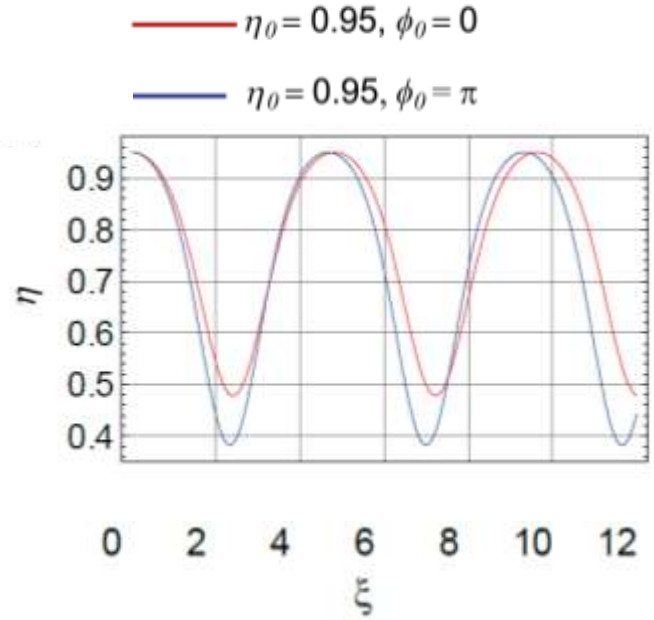
We consider the parametric representation of ideal four wave mixing (FWM) in optical fiber. This is the canonical process underlying frequency comb generation, modulation instability etc.

Nonlinear Schrödinger equation

$$i\frac{\partial A}{\partial \xi} + \frac{1}{2}\frac{\partial^2 A}{\partial \tau^2} + |A|^2 A = 0$$




ODEs (to discover) for relative amplitude and phase

$$\frac{d\eta}{d\xi} = 2\eta^2 \sin\phi - 2\eta\sin\phi$$

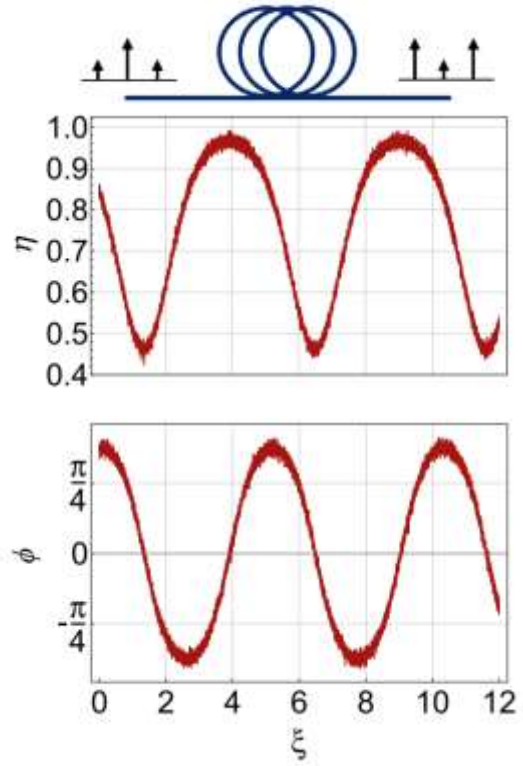
$$\frac{d\phi}{d\xi} = -(\Omega^2 + 1) - 2\cos\phi + 3\eta + 4\eta\cos\phi$$

$$\eta = \frac{|A_0(\xi)|^2}{P_0} = \frac{|A_0(\xi)|^2}{|A_{-1}(\xi)|^2 + |A_0(\xi)|^2 + |A_1(\xi)|^2}$$

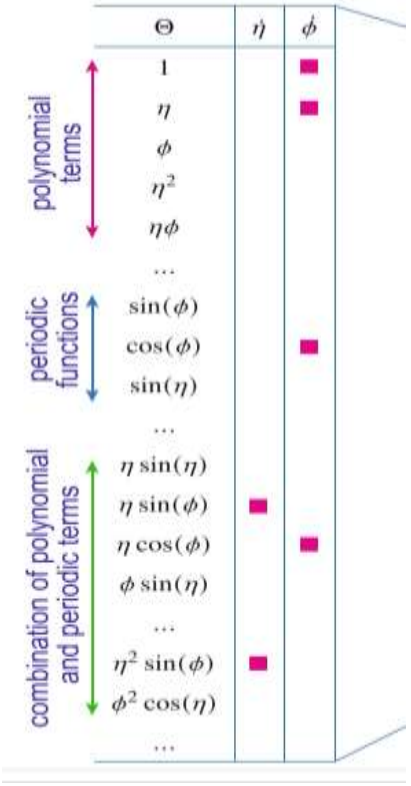
$$\phi = \arg\left[A_{-1}(\xi)\right] - 2\arg\left[A_0(\xi)\right] + \arg\left[A_1(\xi)\right]$$

# Add noise and compute statistics

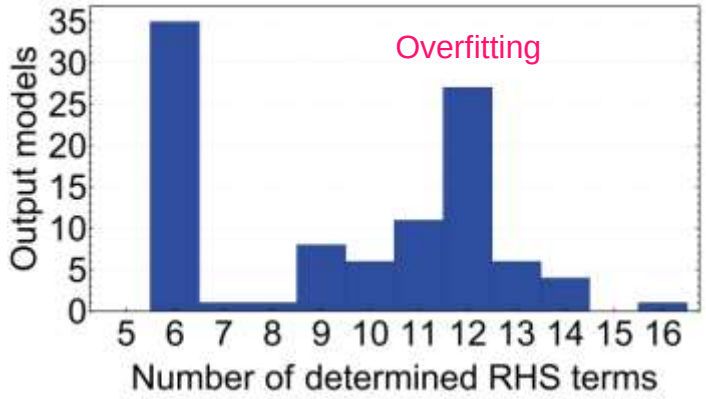
**Input:** 100 data sequences of $\eta(\xi)$ and $\phi(\xi)$ for different initial conditions and 5% noise




**Regression:** fit data with models constructed from up to 32 RHS terms



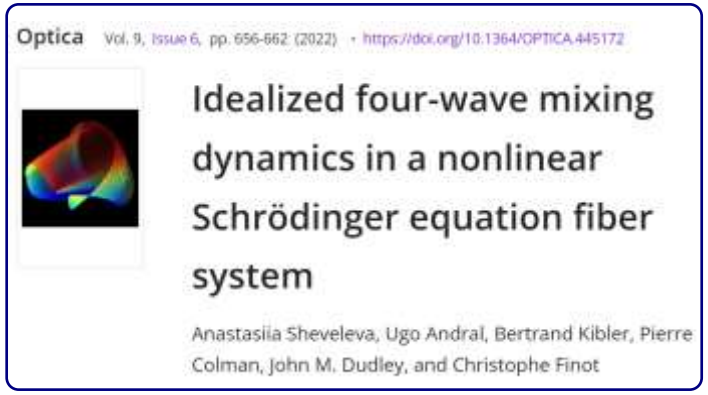

**Results**: the minimally-fitted model agrees with theory

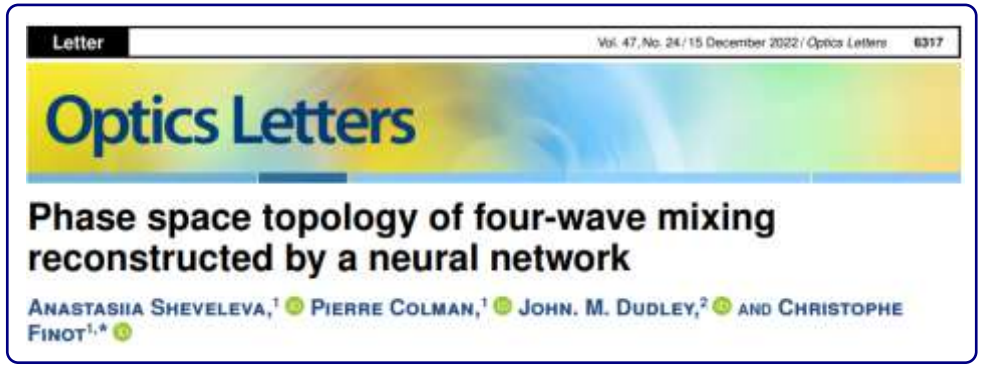


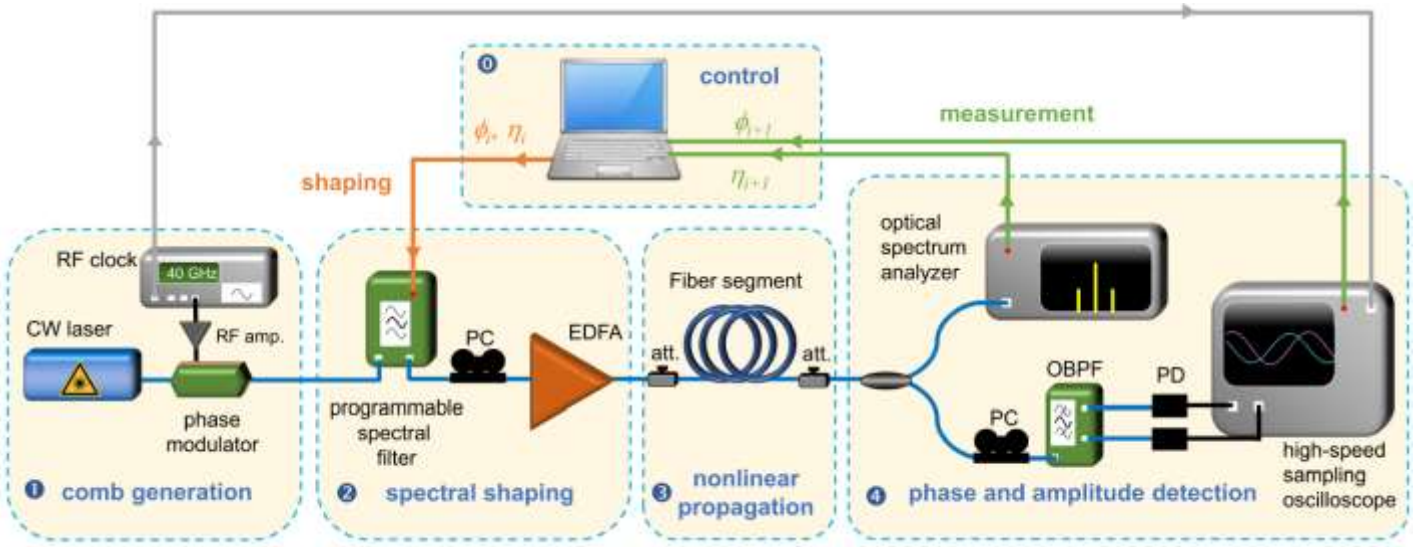

| RHS terms | coefs. vals. | estimated mean | estimated error |
|---|---|---|---|
| $\eta^2 \sin\phi$ | −2 | −1.9978 | $\pm 2.8 \times 10^{-3}$ |
| $\eta \sin\phi$ | 2 | 1.9971 | $\pm 3.8 \times 10^{-3}$ |
| $\eta \cos\phi$ | 4 | 3.9950 | $\pm 4.9 \times 10^{-3}$ |
| 1 | −3 | −2.9989 | $\pm 1.5 \times 10^{-3}$ |
| $\eta$ | 3 | 2.9973 | $\pm 3.4 \times 10^{-3}$ |
| $\cos\phi$ | −2 | −1.9977 | $\pm 2.4 \times 10^{-3}$ |

With noisy data, SINDy generally returns a range of different models with different numbers of terms, but usually the term with the smallest number of terms is the parsimonious model we seek

# We have tried SINDy on experimental data




However SINDy fails with this particular setup. The sideband truncation is too large a deviation from the ideal case.

But this leads to new insights into conservation laws in FWM: Sheveleva *et al.* Annalen der Physik, **536**, 2300489, 2024.

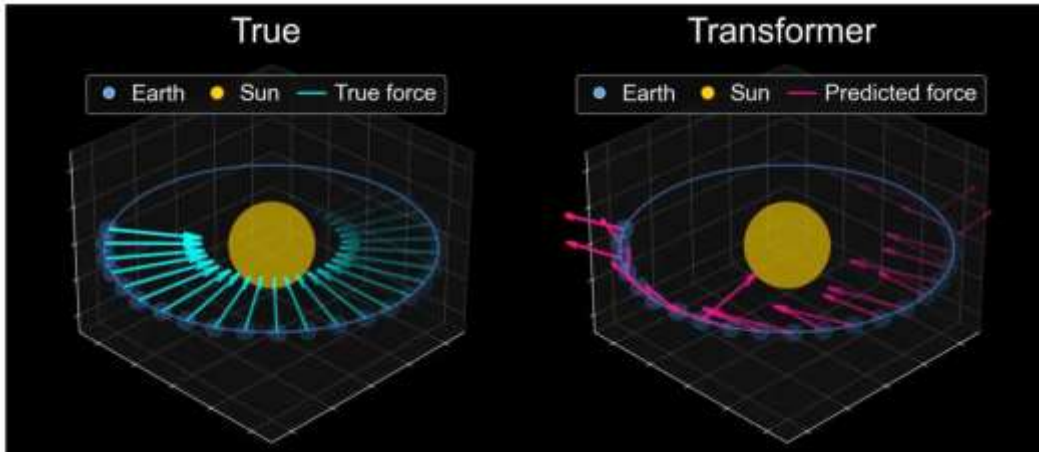
# Is AI really creative? What is creativity in science anyway?

The Algorithmic Bridge

## Harvard and MIT Study: AI Models Are Not Ready to Make Scientific Discoveries

AI can predict the sun will rise again tomorrow, but it can't tell you why

ALBERTO ROMERO
JUL 15, 2025




arXiv:2507.06952v4 [cs.LG] 27 Dec 2025

**What Has a Foundation Model Found?**
**Using Inductive Bias to Probe for World Models**

Keyon Vafa [1] Peter G. Chang [2] Ashesh Rambachan [2] Sendhil Mullainathan [2]

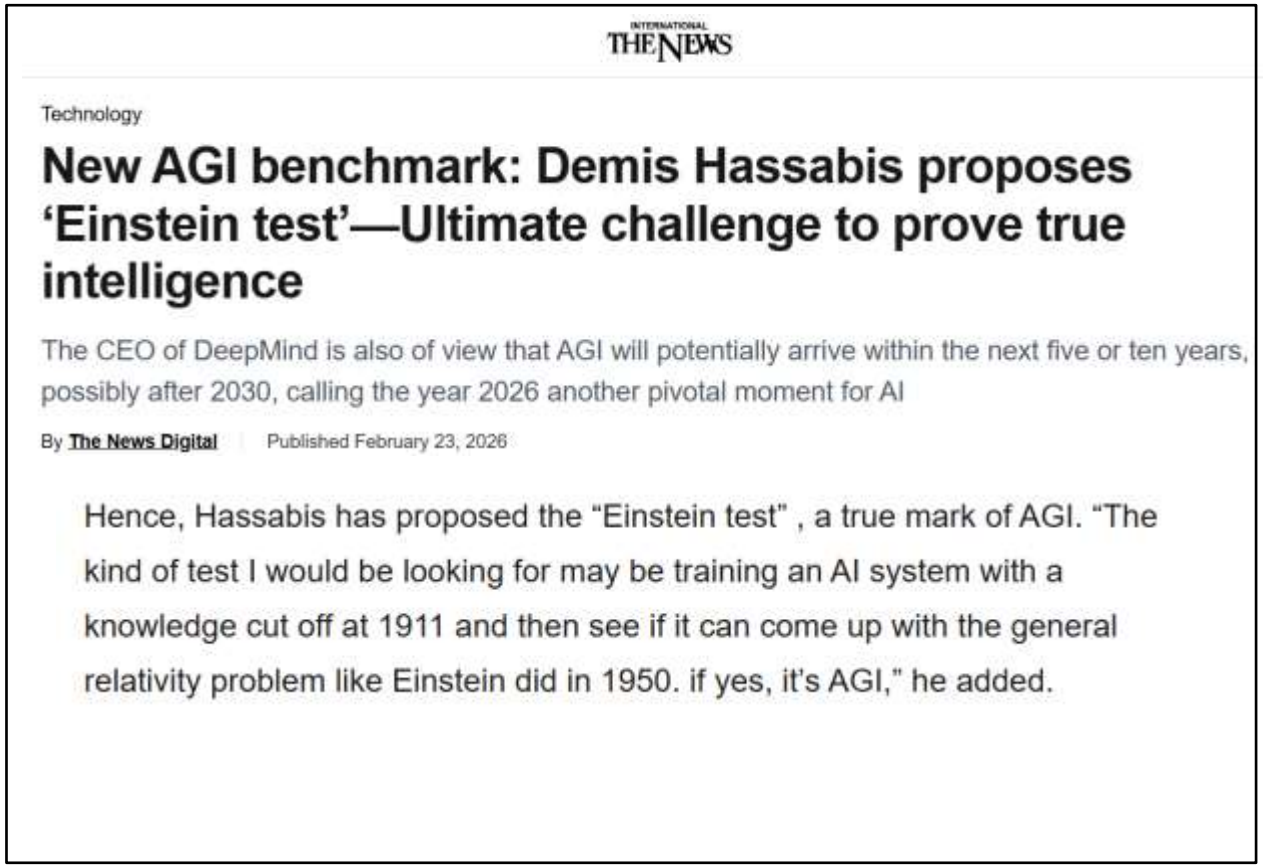
True force law (Newton)

$$F \propto \frac{m_1 m_2}{r^2}$$

Recovered force law (transformer)

$$F \propto \left(\sin\left(\frac{1}{\sin(r-0.24)}\right) + 1.45\right) * \frac{1}{\frac{1}{r^2} + m_2}$$

# Is AI really creative?  What is creativity in science anyway?



INTERNATIONAL THE NEWS

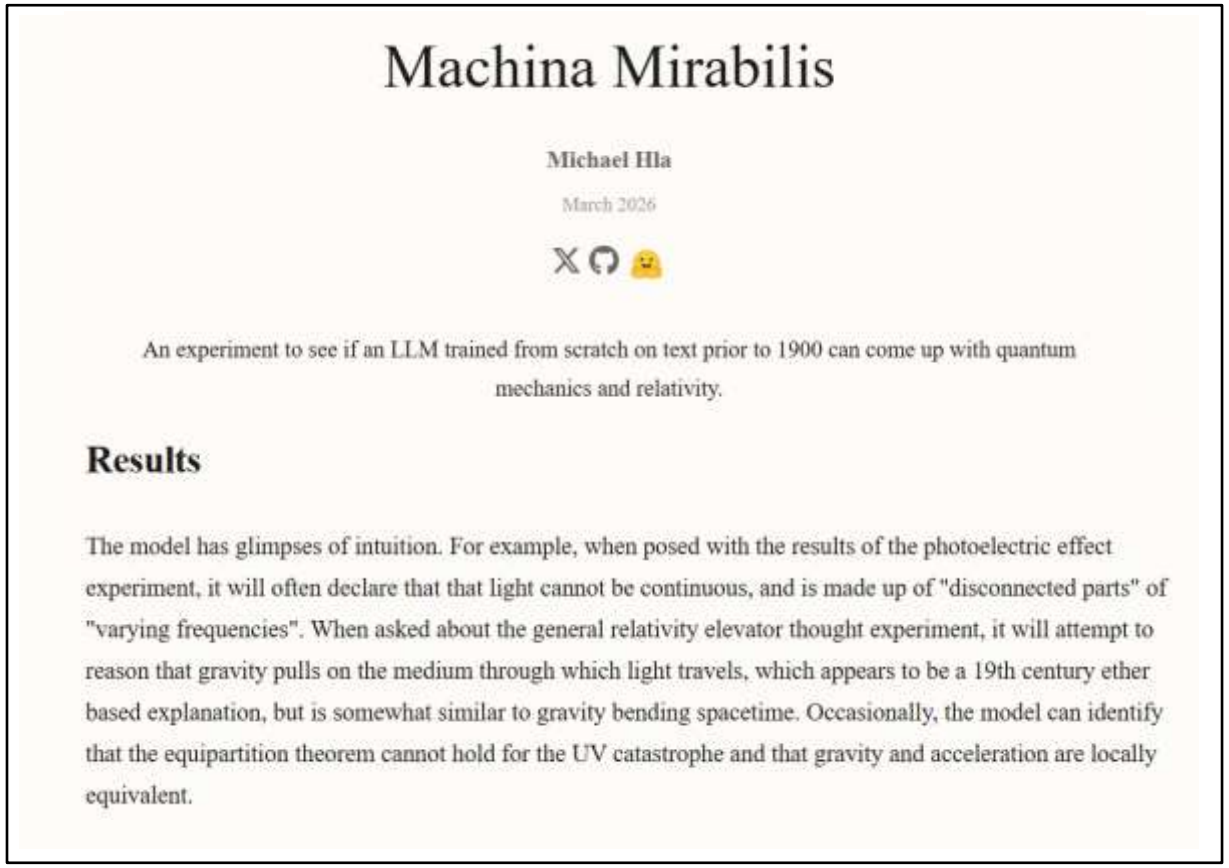
Technology

## New AGI benchmark: Demis Hassabis proposes ‘Einstein test’—Ultimate challenge to prove true intelligence

The CEO of DeepMind is also of view that AGI will potentially arrive within the next five or ten years, possibly after 2030, calling the year 2026 another pivotal moment for AI

By The News Digital | Published February 23, 2026

Hence, Hassabis has proposed the “Einstein test” , a true mark of AGI. “The kind of test I would be looking for may be training an AI system with a knowledge cut off at 1911 and then see if it can come up with the general relativity problem like Einstein did in 1950. if yes, it’s AGI,” he added.



# Machina Mirabilis

Michael Hla

March 2026

An experiment to see if an LLM trained from scratch on text prior to 1900 can come up with quantum mechanics and relativity.

## Results

The model has glimpses of intuition. For example, when posed with the results of the photoelectric effect experiment, it will often declare that that light cannot be continuous, and is made up of "disconnected parts" of "varying frequencies". When asked about the general relativity elevator thought experiment, it will attempt to reason that gravity pulls on the medium through which light travels, which appears to be a 19th century ether based explanation, but is somewhat similar to gravity bending spacetime. Occasionally, the model can identify that the equipartition theorem cannot hold for the UV catastrophe and that gravity and acceleration are locally equivalent.

# Automating physical intuition through dominant balance search

**Concept of dominant balance**

For any differential equation, the sum of terms will always equal zero, but different subsets of terms will dominate the equality in different spatio-temporal regions.

NATURE COMMUNICATIONS | (2021)12:1016 |

Learning dominant physical processes with data-driven balance models

Jared L. Callaham[1✉], James V. Koch[2], Bingni W. Brunton[3], J. Nathan Kutz[4] & Steven L. Brunton[1]

**The question**

Can we automatically identify the most significant terms that balance the equation in different points/regions $(\xi,\tau)$.

$$\sum_{i-1}^{K} f_i(\psi, \psi_\xi, \psi_\tau \cdots, \psi^2, \psi\psi_\xi, \psi\psi_\tau, \ldots, \psi_{\xi\xi}, \psi_{\tau\tau}, \ldots) = 0$$

$$i\psi_\xi + \psi_{\tau\tau} + i\delta\psi_{\tau\tau\tau} + |\psi|^2\psi + \rho\psi\,(h_R * |\psi|^2) = 0$$

GVD — Group velocity dispersion

TOD — Third-order dispersion

SPM — Self phase modulation

Raman

Dominant balances manifest as clusters of reduced variance in the multidimensional **equation space,** aligned along low dimension subspaces (e.g. planes where only a few terms cancel each other).

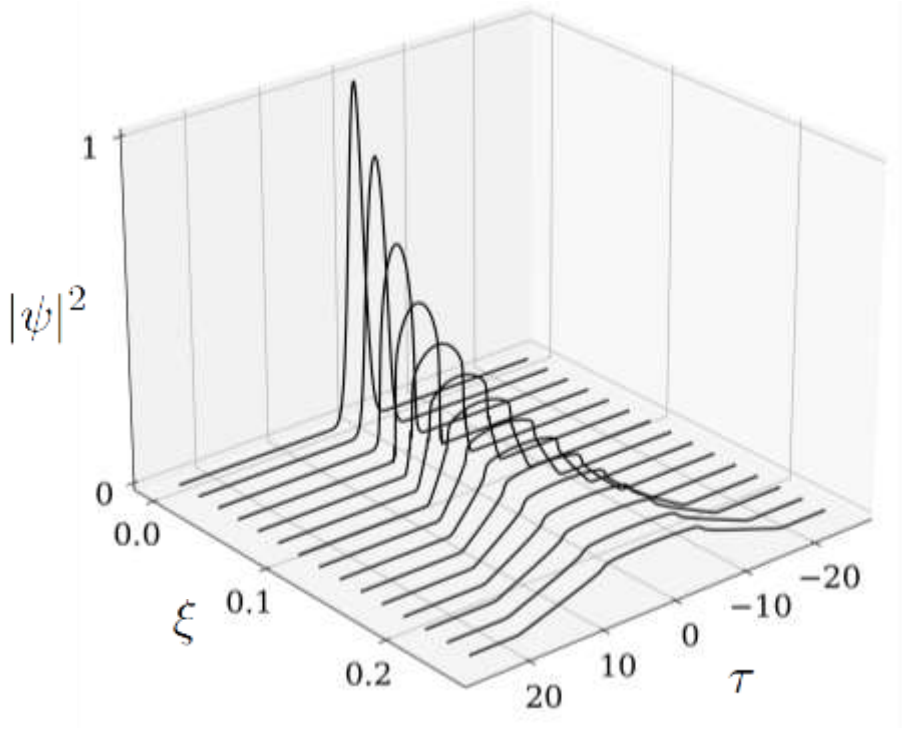
# An example from nonlinear fibre optics – optical wavebreaking

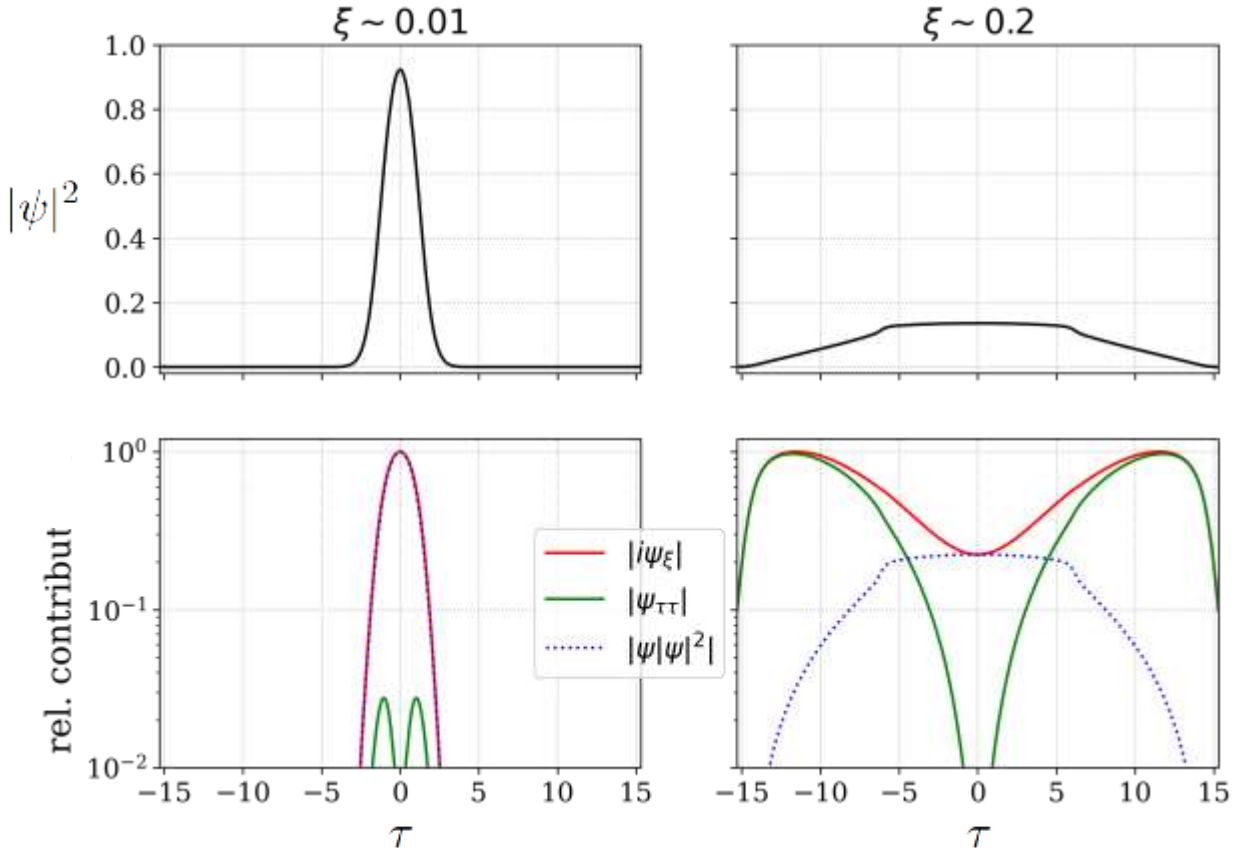
Nonlinear Schrödinger Equation (NLSE)

$$i\,\psi_\xi - \psi_{\tau\tau} + |\psi|^2\psi = 0$$

$$\psi(0,\tau) = N\,\exp\left(-\tau^2/4\right)$$

$$N = T_0\sqrt{\gamma P_0/|\beta_2|} = 30$$




The NLSE equality is satisfied by different subsets of terms at different $\xi$,$\tau$.

W. J. Tomlinson, R. H. Stolen & A. M. Johnson, *Optics Letters* **10**, 457–459 (1985)

# Balance map of optical wavebreaking

$$i\,\psi_\xi - \psi_{\tau\tau} + |\psi|^2\psi = 0$$

$$\psi(0,\tau) = N\,\exp\left(-\tau^2/4\right)$$

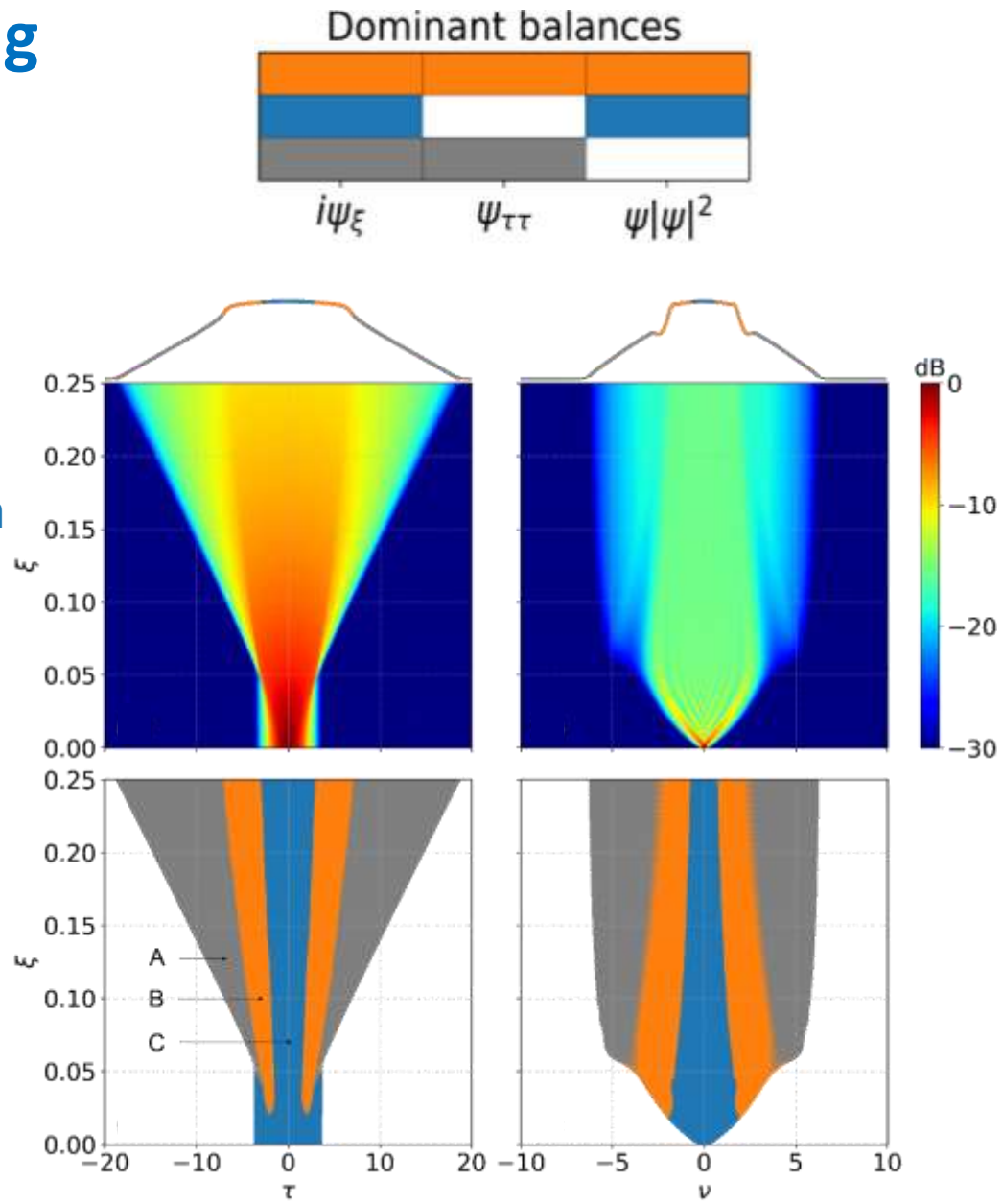
$$N = T_0\sqrt{\gamma P_0/|\beta_2|} = 30$$

Optical wavebreaking occurs with high power propagation in the normal dispersion regime of an optical fibre

A. Dispersion dominant in temporal wings

B. NLSE dominates the transition regime

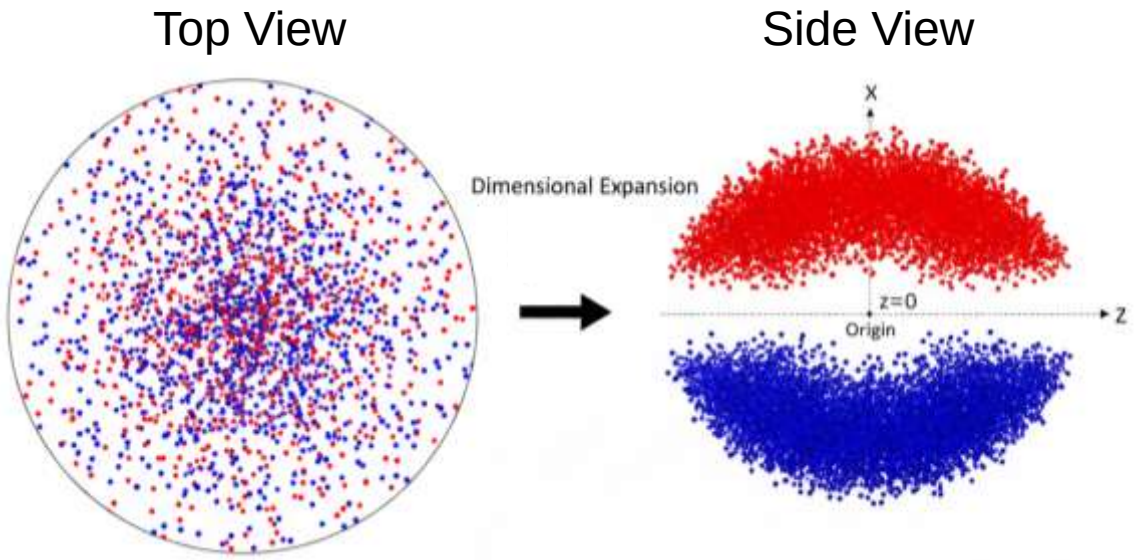
C. Nonlinearity dominates temporal centre



# Supercontinuum computing

# Nonlinear optics for unconventional computing

One approach to exploit nonlinearity for computation is the **extreme learning machine** where low-dimensional information is transformed to a high-dimensional space such that indistinguishable features in the original input space become separable and we can use gradient-free regression to classify.



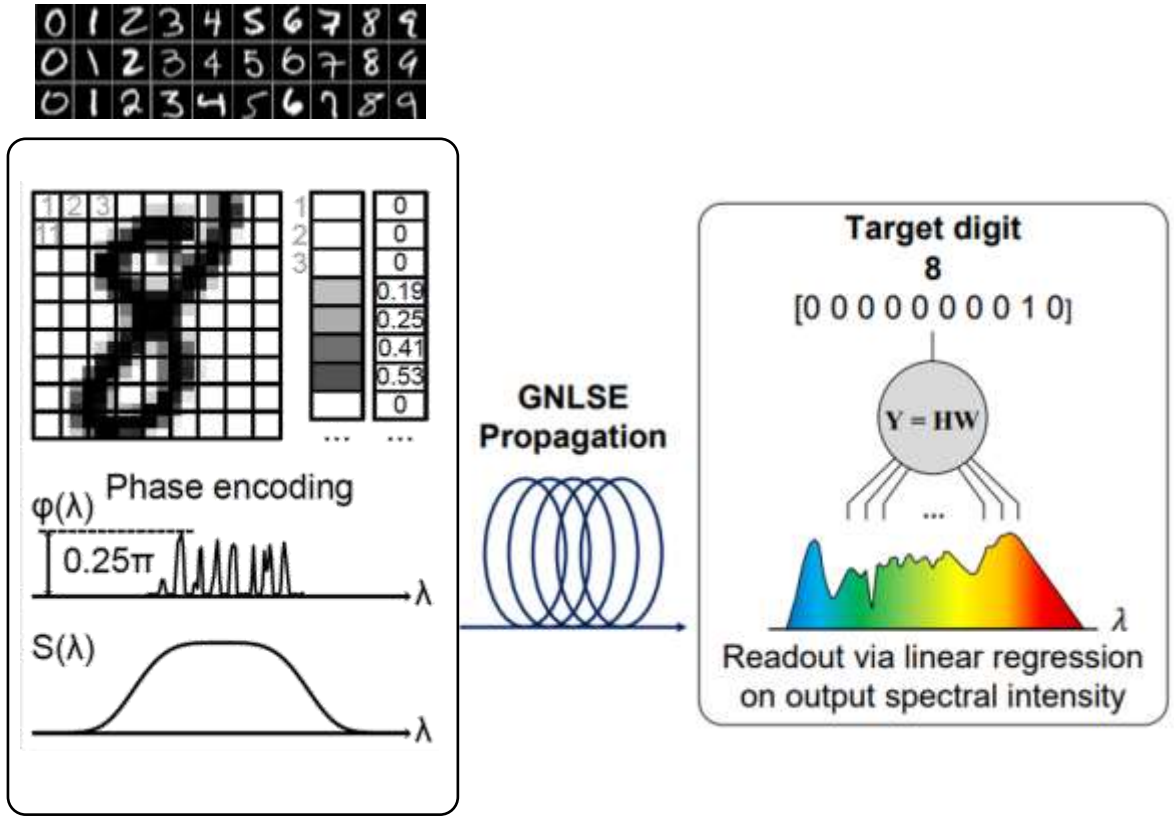

IEEE TRANSACTIONS ON ELECTRONIC COMPUTERS

Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition

THOMAS M. COVER

**Thomas M. Cover** (1938–2012)

# The supercontinuum extreme learning machine




**Encoding:** Image pixel data is mapped to e.g. 100 wavelengths on a 20 nm pulse spectrum

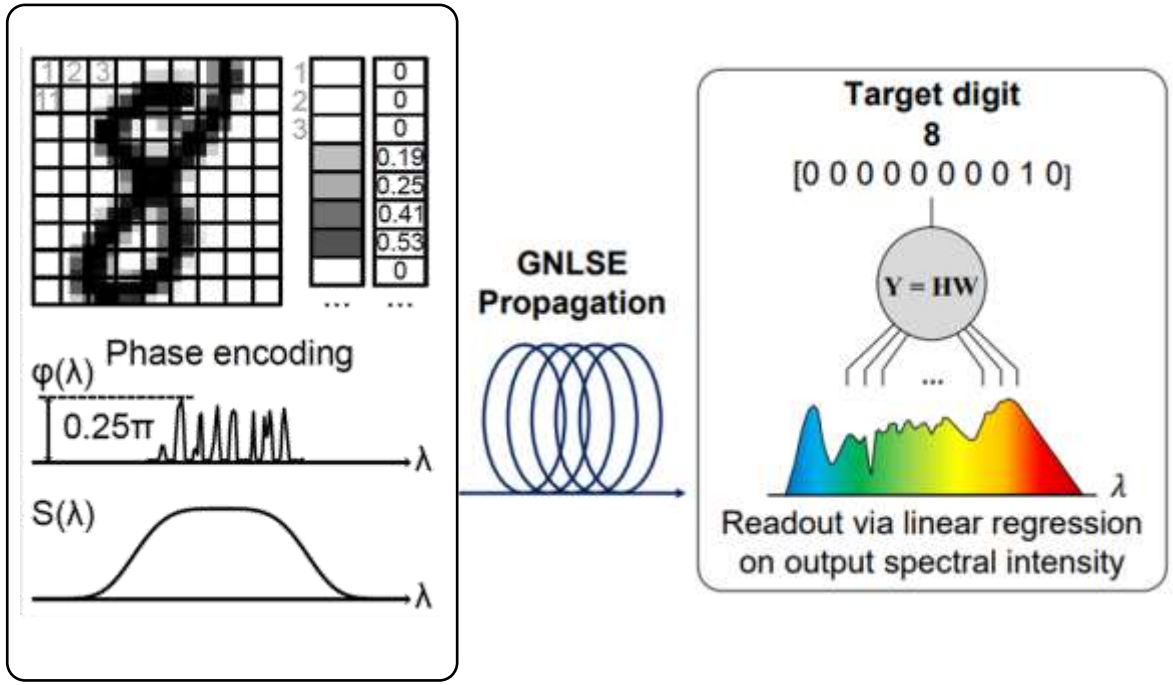
**Dimensional Expansion:** Nonlinear propagation expands the input pulse to 1000 nm bandwidth.

**Feature mapping:** Different handwritten digits produce similar structure on the broadening output spectra, allowing these broadened spectra to be used to categorize the data.

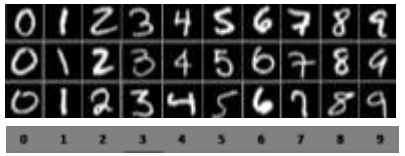
This is an example of projecting simple data into a much **higher-dimensional space** where patterns are easier to identify.

# The supercontinuum extreme learning machine

We have developed a **simulation pipeline** to determine the theoretical limits of this approach




$\mathbf{X}$ Input: Hand drawn images



$\mathbf{Y}$ Output: Corresponding digits

$\mathbf{H} = f_{\mathrm{NL}}(\mathbf{X})$ Compute the nonlinear transformation of $\mathbf{X}$ to form the "hidden layer"

$\mathbf{Y}^{\mathrm{T}} \approx \mathbf{H}\mathbf{W}^{\mathrm{out}}$ Based on this computed $\mathbf{H}$ and the known digits $\mathbf{Y}^{\mathrm{T}}$ we can determine the weight matrix $\mathbf{W}^{\mathrm{out}}$ in a single step,

$\mathbf{W}^{\mathrm{out}} \approx \mathbf{H}^{\dagger}\,\mathbf{Y}^{\mathrm{T}}$ $\mathbf{H}^{\dagger}$ is the pseudo-inverse matrix

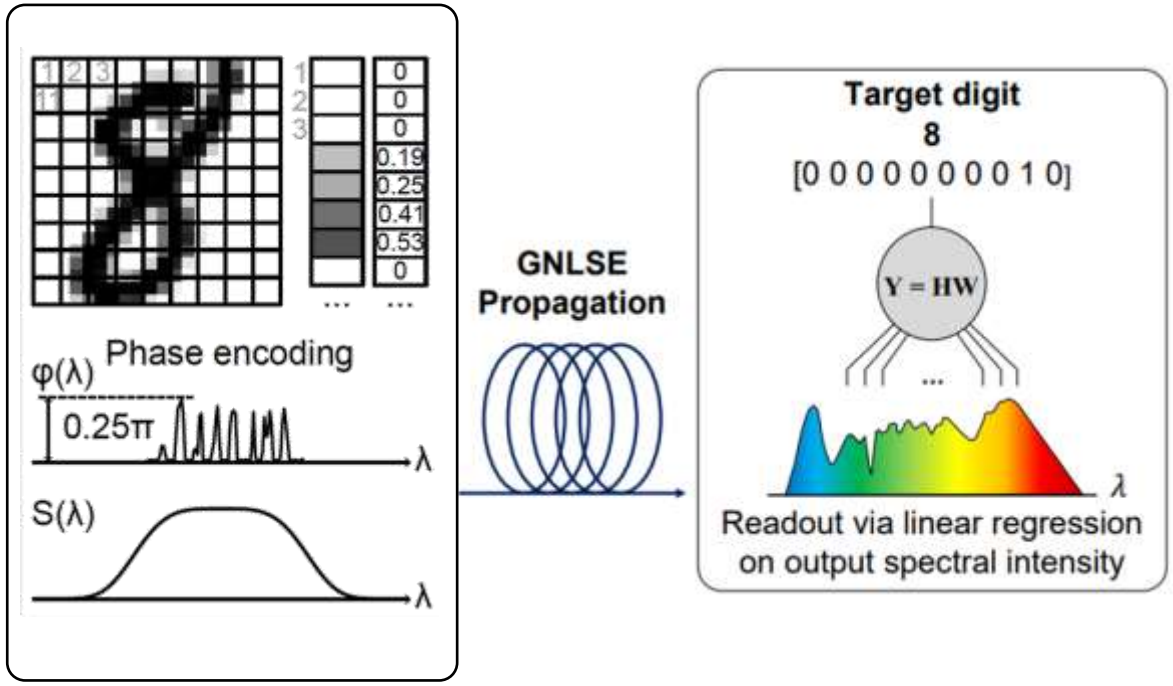
$$i\frac{\partial A}{\partial z} - \frac{1}{2}\beta_2\frac{\partial^2 A}{\partial T^2} - \frac{i}{6}\beta_3\frac{\partial^3 A}{\partial T^3} + \frac{1}{24}\beta_4\frac{\partial^4 A}{\partial T^4} + \gamma\left(1 + i\omega_0\frac{\partial}{\partial T}\right)\left(A\left[R * |A|^2\right]\right) = 0$$

$$N = (\gamma P_0 T_0^2/|\beta_2|)^{1/2} \qquad \Delta\tau \sim 182 \text{ fs FWHM}$$

We determine accuracy by applying $\mathbf{W}^{\mathrm{out}}$ to a new set of data $\mathbf{X}$ and predicting the corresponding digits

# The supercontinuum extreme learning machine

We have developed a **simulation pipeline** to determine the theoretical limits of this approach




60000 “training” images

$$\mathbf{X} = \quad 60000 \times 100$$

$$\mathbf{Y}^{\mathrm{T}} = \begin{bmatrix} 0\,0\,0\,0\,0\,0\,0\,0\,1\,0 \\ 0\,1\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0\,0\,0\,0\,0 \\ \\ 0\,1\,0\,0\,0\,0\,0\,0\,0\,0 \end{bmatrix} \quad 60000 \times 10$$

The hidden layer matrix consists of the generated supercontinuum spectrum for each phase encoded image, after readout over $K$ wavelength bands

$$\mathbf{H} = \quad 60000 \times K$$

$$\mathbf{W}^{\mathrm{out}} \approx \mathbf{H}^{\dagger}\,\mathbf{Y}^{\mathrm{T}} = \quad K \times 60000 \quad 60000 \times 10 \quad K \times 10$$

$$i\frac{\partial A}{\partial z} - \frac{1}{2}\beta_2\frac{\partial^2 A}{\partial T^2} - \frac{i}{6}\beta_3\frac{\partial^3 A}{\partial T^3} + \frac{1}{24}\beta_4\frac{\partial^4 A}{\partial T^4} + \gamma\left(1 + i\omega_0\frac{\partial}{\partial T}\right)\left(A\left[R * |A|^2\right]\right) = 0$$

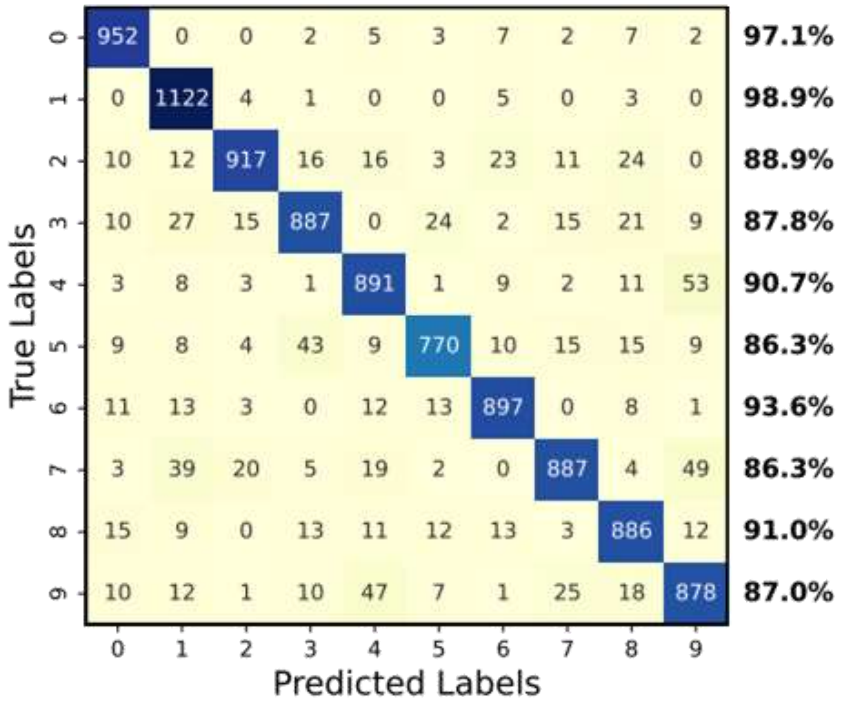
$$N = (\gamma P_0 T_0^2/|\beta_2|)^{1/2} \qquad \Delta\tau \sim 182 \text{ fs FWHM}$$

We determine accuracy by applying $\mathbf{W}^{\mathrm{out}}$ to a new set of data $\mathbf{X}$ and predicting the corresponding digits

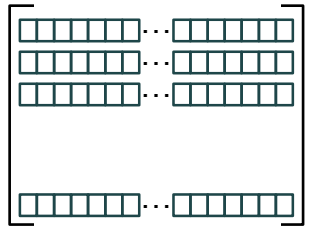
# The supercontinuum extreme learning machine

We have developed a **simulation pipeline** to determine the theoretical limits of this approach.

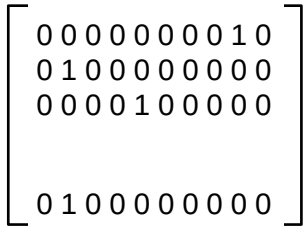
Test accuracy 90.9%




10000 “test” images

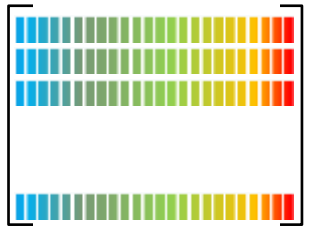
$\mathbf{X} =$ 

10000 x 100

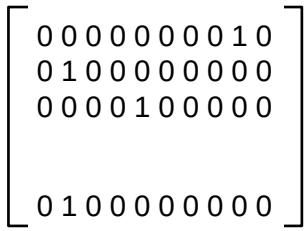
$\mathbf{Y} =$ 


10000 x 10

The hidden layer matrix consists of the generated supercontinuum spectrum for each phase encoded image, after readout over $K$ wavelength bands

$\mathbf{H} =$ 

10000 x K

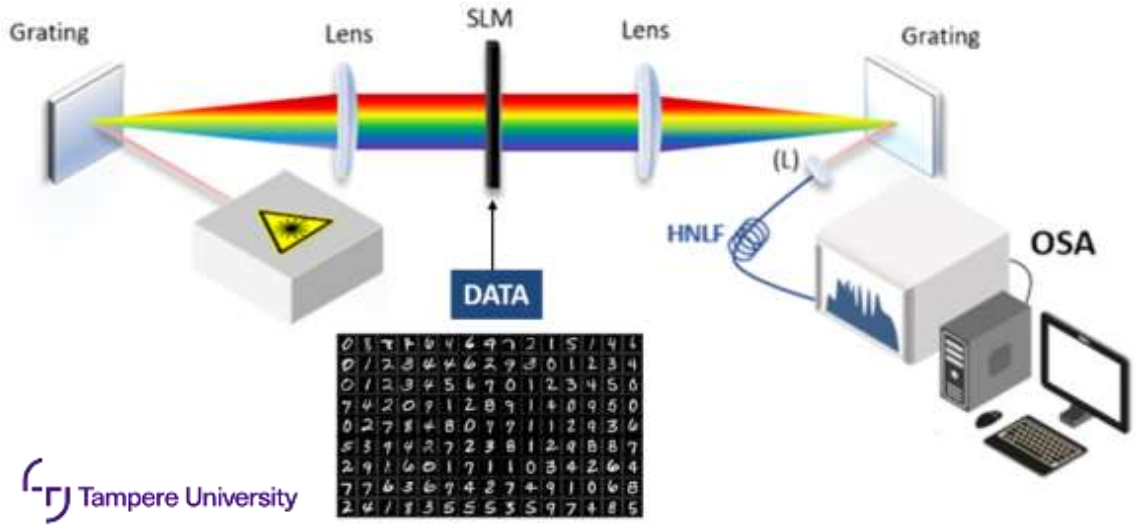
$\mathbf{Y}^{\text{pred}} \approx \mathbf{H}\mathbf{W}^{\text{out}}$ (10000 x K, K x 10) $\xrightarrow[\text{one\_hot}]{\text{argmax}}$ 


10000 x 10

We determine accuracy by applying $\mathbf{W}^{\text{out}}$ to a new set of data $\mathbf{X}$ and test the predictions of the corresponding digits

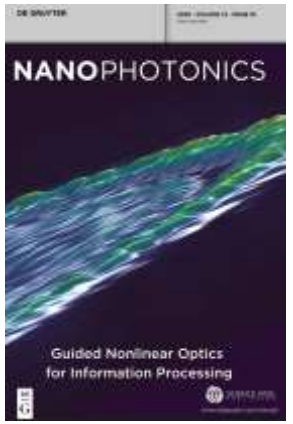
# The supercontinuum extreme learning machine

It is straightforward to compare results for normal and anomalous dispersion regime propagation over a wide range of parameters (power, length, encoding, readout)




Although detailed differences with **published experiments** preclude a direct comparison, the accuracies reported are close to predictions.

| | |
|---|---|
| **Anomalous Dispersion** | |
| Simulations | 90.9% |
| Fischer *et al.* (2023) | 86.7% |
| Saeed *et al.* (2025) | 87.3% |
| **Normal Dispersion** | |
| Simulations | 92.9% |
| Saeed *et al.* (2025) | 89.3% |
| Hary *et al.* (2025) | 88.0% |
| Linear benchmark | ~ 85.5% |




Input quantum noise penalty ~ 2-3 %

# Conclusions

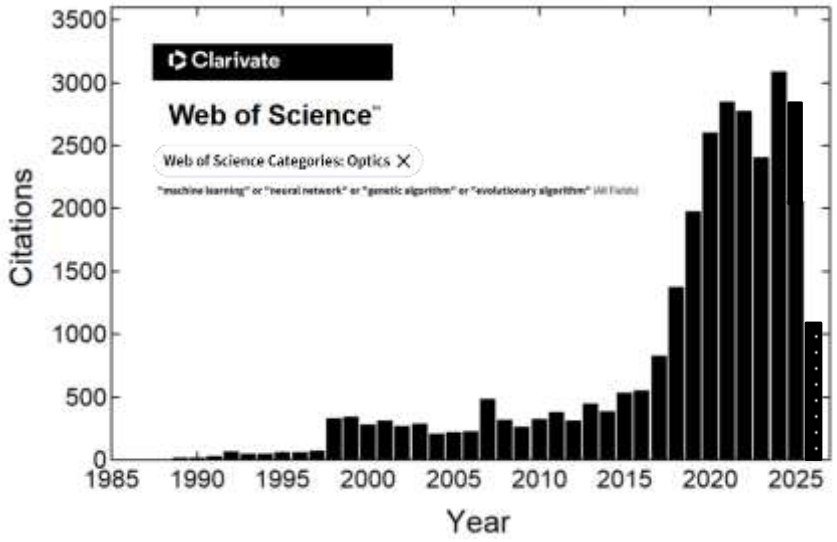
Most of the work in theoretical/numerical methods has really only scratched the surface.

- The obvious direction for SINDy is application to partial differential equations in optics.
- Dominant balance methods are beginning to show value in adding physical interpretation.

Proof of principle source optimization is complete and is moving to practical sources.

Fibre systems are convenient for the study of nonlinear kernel methods in optical computing, but are best seen as testbeds for approaches that will be implemented in other platforms.

# Where are we now?


Gartner Hype Cycle


## Some helpful context: Feynman's problem with the worldview of quantum mechanics

I cannot define the real problem, therefore I suspect there's no real problem, but I'm not sure there's no real problem. So that's why I like to investigate.

Can I learn anything from asking this question about computers?

R. P. Feynman, *Simulating Physics with Computers.* Int. J. Theor. Phys. **21** 6/7 (1982)

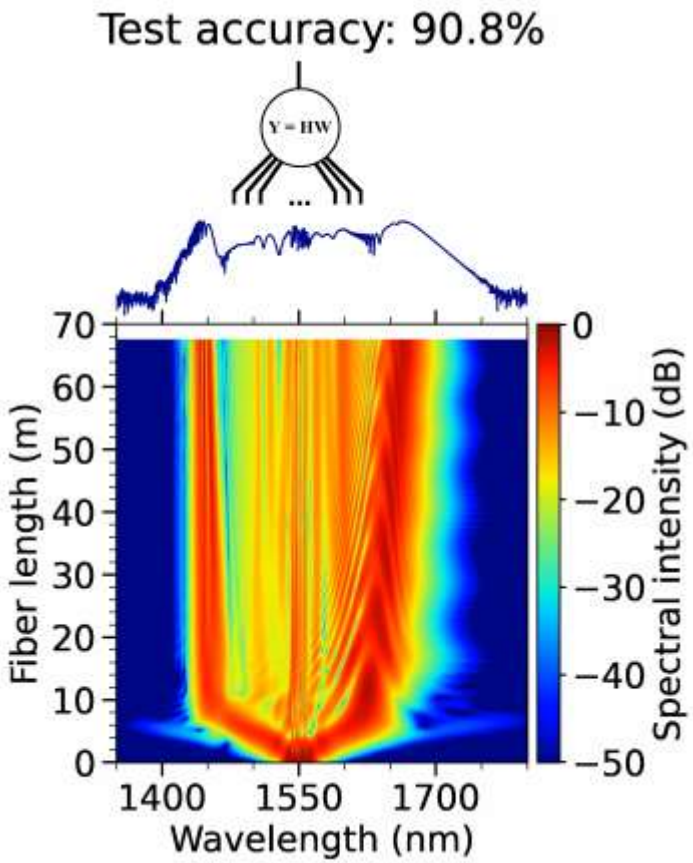
# One of the most famous photos in the history of science


Physics of Computation Conference Endicott House MIT May 6-8, 1981

1 Freeman Dyson
2 Gregory Chaitin
3 James Crutchfield
4 Norman Packard
5 Panos Ligomenides
6 Jerome Rothstein
7 Carl Hewitt
8 Norman Hardy
9 Edward Fredkin
10 Tom Toffoli
11 Rolf Landauer
12 John Wheeler
13 Frederick Kantor
14 David Leinweber
15 Konrad Zuse
16 Bernard Zeigler
17 Carl Adam Petri
18 Anatol Holt
19 Roland Vollmar
20 Hans Bremerman
21 Donald Greenspan
22 Markus Buettiker
23 Otto Floberth
24 Robert Lewis
25 Robert Suaya
26 Stan Kugell
27 Bill Gosper
28 Lutz Priese
39 Madhu Gupta
30 Paul Benioff
31 Hans Moravec
32 Ian Richards
33 Marian Pour-El
34 Danny Hillis
35 Arthur Burks
36 John Cocke
37 George Michaels
38 Richard Feynman
39 Laurie Lingham
40 Thiagarajan
41 ?
42 Gerard Vichniac
43 Leonid Levin
44 Lev Levitin
45 Peter Gacs
46 Dan Greenberger

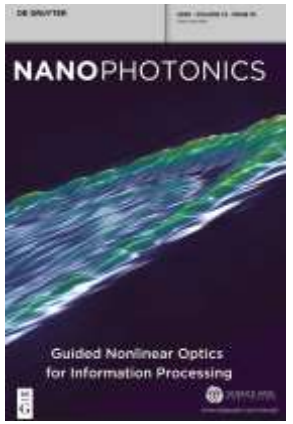
# The supercontinuum extreme learning machine

It is straightforward to compare results for normal and anomalous dispersion regime propagation over a wide range of parameters (power, length, encoding, readout)



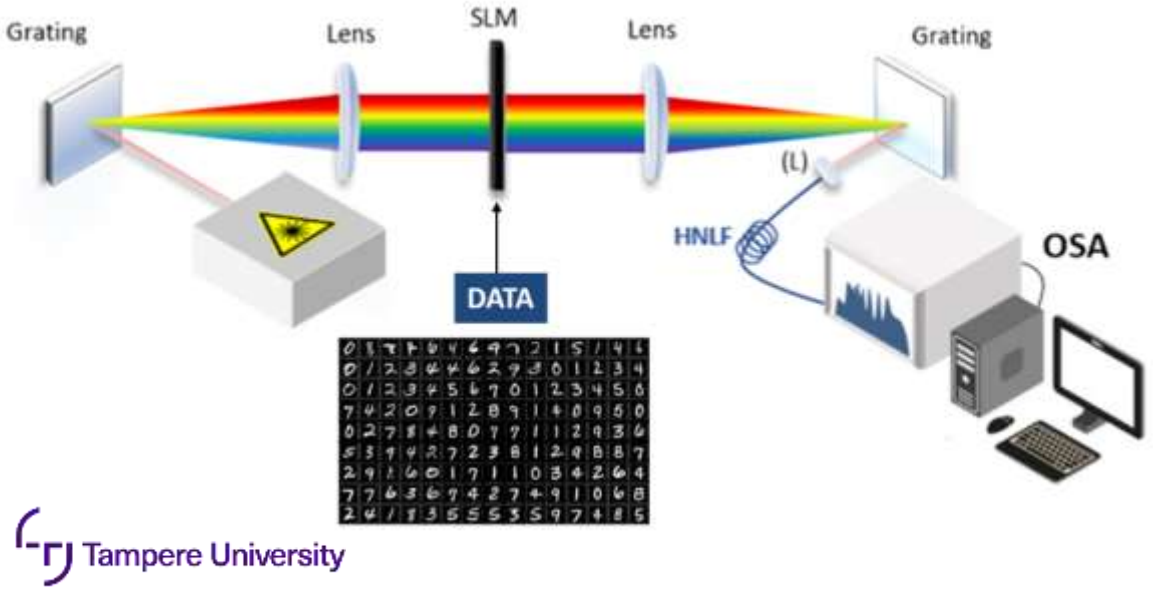

Although detailed differences with **published experiments** preclude a direct comparison, the accuracies reported are close to predictions.

| | |
|---|---|
| **Anomalous Dispersion** | |
| Simulations | 90.9% |
| Fischer *et al.* (2023) | 86.7% |
| Saeed *et al.* (2025) | 87.3% |
| **Normal Dispersion** | |
| Simulations | 92.9% |
| Saeed *et al.* (2025) | 89.3% |
| Hary *et al.* (2025) | 88.0% |
| Linear benchmark | ~ 85.5% |



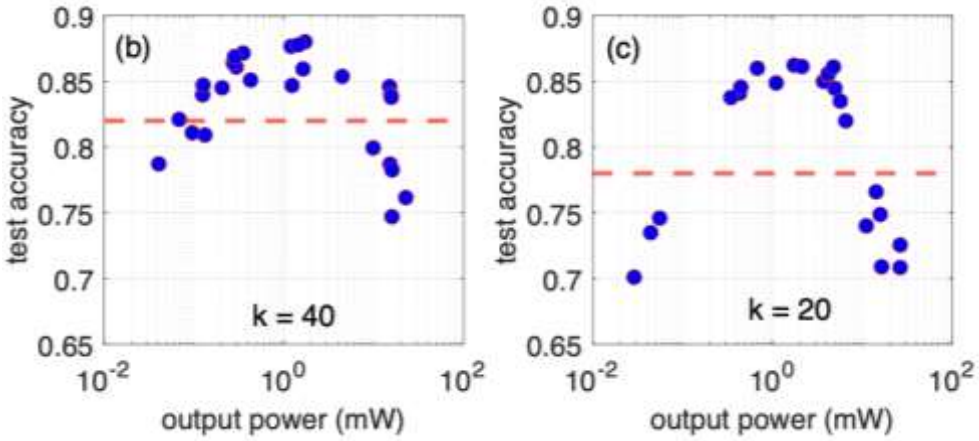

Input quantum noise penalty ~ 2-3 %

# Experimental implementation




Tampere University

Experimental studies studied test accuracy as a function of input PCA and power




**NKT Photonics ORIGAMI source**

$\lambda_0$ = 1559 nm

$T_R$ = 40.9 MHz

$\Delta\tau$ ~ 200 fs (14 nm FWHM)

$E_{in}$ ~ 2 nJ

Different types of HNLF exploring different dispersion regimes

**Principles and Metrics of Extreme Learning Machines Using a Highly Nonlinear Fiber**

Mathilde Hary,[1,2] Daniel Brunner,[2] Lev Leybov,[1] Piotr Ryczkowski,[1] John M. Dudley,[3] and Goëry Genty[1]

# (Aside: a lesson for today's research culture)

1814
Fraunhofer sees dark lines in the solar spectrum

1885
Balmer finds a formula for the spectral lines of hydrogen

$$\frac{1}{\lambda} = R\left(\frac{1}{2^2} - \frac{1}{n^2}\right)$$

$n = 3, 4, 5, 6, \ldots$

$R = 10973731.57\,\mathrm{m}^{-1}$

1913
The Bohr model

1859
Bunsen & Kirchoff see bright lines at the same position as Fraunhofer's dark lines

1900-1905
Planck and Einstein introduce quanta of energy exchange and quanta of light energy

99 years between the first measurements of Fraunhofer lines and the Bohr model

27 years between the data-driven derivation of the Balmer formula and the Bohr model

Science takes time and needs long-term support

# Machine learning is impacting all of science

The tools of artificial intelligence and machine learning are having tremendous impact across science, including of course in ultrafast and nonlinear optics and photonics

**Nobel Prizes awarded to AI-Pioneers:**

**The Deep link between Computing and Physics (and Chemistry)**

Daniel BRUNNER[1,*], John M. DUDLEY[2], Demetri PSALTIS[2]

[1] FEMTO-ST/Optics Department, UMR CNRS 6174, Université Franche-Comté, 25030 Besançon Cedex, France

[2] Optics Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

* daniel.brunner@femto-st.fr

**The 2024 Nobel Prize in Physics has been awarded to John J. Hopfield and Geoffrey E. Hinton for their pioneering contributions to harnessing principles from physics to establish foundational methods in machine learning. Their work catalysed ground-breaking computing concepts, establishing the basis for unconventional computing architectures that directly exploit the physics of their underlying hardware. Today, these novel paradigms promise to drive next-generation hardware with enhanced performance and efficiency, while advanced neural network architectures open doors to transformative scientific discoveries. In this review, we outline the broader context of their contributions to unconventional computing and emphasize the integration of physics-based concepts in modern machine learning architectures.**

https://doi.org/10.1051/photon/202212812

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The human brain, a biological marvel, has inspired the development of physical computers with extraordinary capabilities. Notably, it is highly energy-efficient, capable of complex problem-solving, supports creativity and can solve certain problems with remarkable reliability. Research into brain function has long had two interwoven goals: to understand the brain itself, and to develop technologies that mirror its capabilities. This dual interest has influenced generations of scientists, including the 2024 Nobel laureates in Physics, Geoffrey Hinton and John Hopfield, whose work has highlighted the profound link between brain-inspired computing and fundamental physical principles [1, 2].

Early computing began with special-purpose devices that leveraged analogue physics to perform calculations, usually with a rather direct mapping of the computational task onto the physical laws of the machines. This approach, though relatively simple to implement, limited such devices in their ability to evaluate/solve/compute/mimic specific functions to the ones embedded in their physical construction. The breakthrough in general-purpose computing emerged with Alan Turing's concepts, which abstracted computation using binary logic [3]. By reducing computation to simple, hence more readily reproducible operations, Turing's model made computing robust against physical imperfections, enabling the development of reliable hardware that operated largely independent of the specifics of the underlying physics.

However, this direct task-mapping approach, linking computational objectives to Turing algorithms, diverges

John J. Hopfield and Geoffrey Hinton © Niklas Elmehed - Nobel Prize Outreach