PHOTONICS-BASED MACHINE LEARNING TO SPEED UP AND SIMPLIFY LABEL-FREE FLOW CYTOMETRY

Alessio Lugnan - 7 September 2021

Promotors: Peter Bienstman, Joni Dambre

GHENT UNIVERSITY

imec
embracing a better life

OUTLINE

Introduction:

- Technique to improve: flow cytometry
- Approach: machine learning and neural networks
- Problem: microparticle classification algorithms limit the speed of flow cytometry
- Solution: a hardware-based machine learning approach

White blood cell hologram classification

Dimensionality expansion with dielectric scatterers

Development of flow cytometry experiment

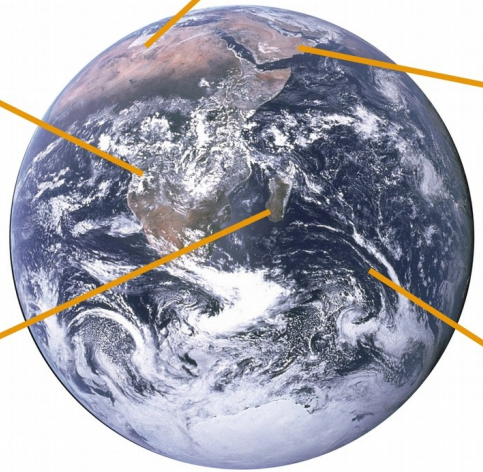
Final experiment results

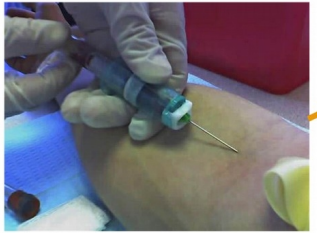
INTRODUCTION

A WORLD OF INTERESTING MICROPARTICLES

Liquids can host huge numbers and varieties of microscopic objects and life forms, for example:

- cells in blood
- microbes in water and food
- pollutants (e.g. microplastics) in water
- plankton in the ocean
- ...

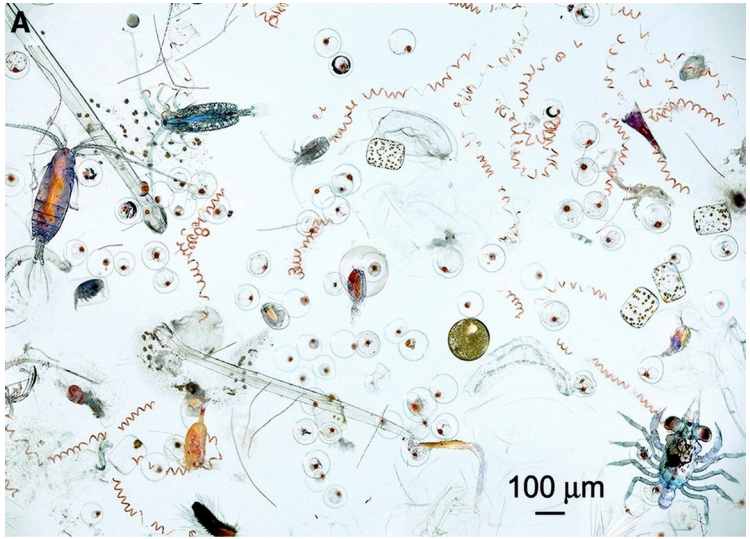




Images from Wikipedia.org

NUMBERS MATTER

Statistical validity of scientific studies or **detection of rare objects** often require a large number of single-object measurements
→ **Flow cytometry** allows to analyse microscopic objects one by one, in a flow at high speed



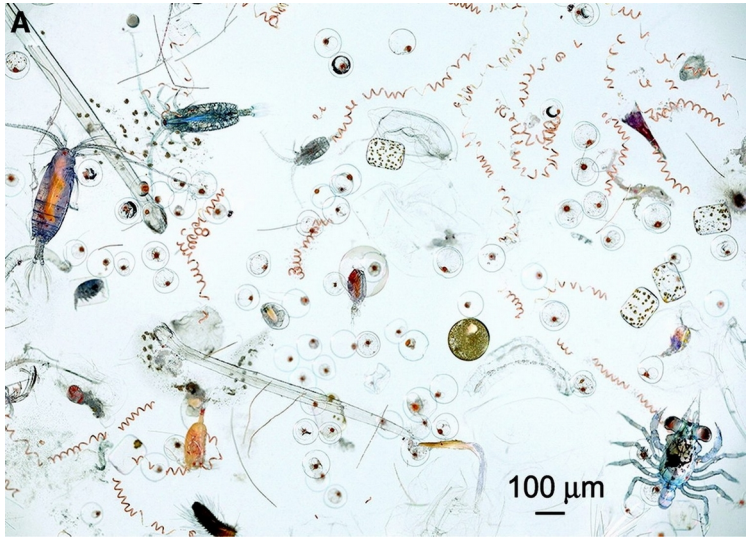
Wikipedia.org

NUMBERS MATTER

Statistical validity of scientific studies or **detection of rare objects** often require a large number of single-object measurements
→ **Flow cytometry** allows to analyse microscopic objects one by one, in a flow at high speed

Applications:

- Biological analysis of heterogeneous cell populations
- Cell sorting, to automatically isolate specific cell types
- Detection of circulating tumor cells in blood
- Blood analysis to monitor immune status
- Monitoring of waterborne microbes for water treatment and reuse
- Bacteria viability in probiotic products
- ...



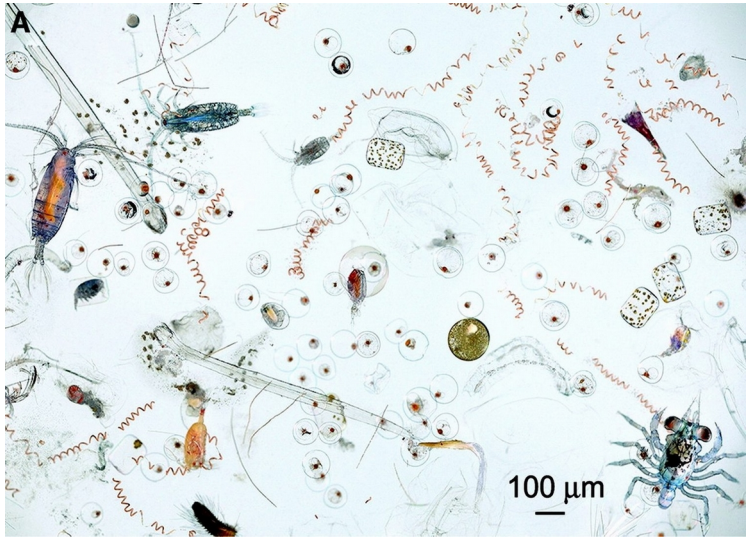
Wikipedia.org

NUMBERS MATTER

Statistical validity of scientific studies or **detection of rare objects** often require a large number of single-object measurements
→ **Flow cytometry** allows to analyse microscopic objects one by one, in a flow at high speed

Applications:

- Biological analysis of heterogeneous cell populations
- Cell sorting, to automatically isolate specific cell types
- Detection of circulating tumor cells in blood
- Blood analysis to monitor immune status
- Monitoring of waterborne microbes for water treatment and reuse
- Bacteria viability in probiotic products
- ...



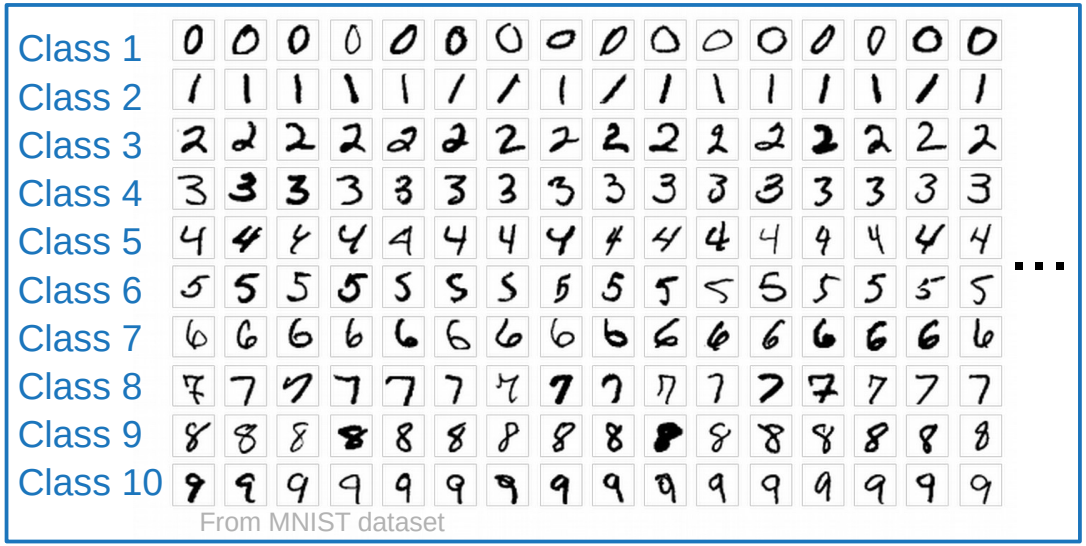
Wikipedia.org

The related scientific community aims to make cytometers more compact, cheap, easy to use and fully automatic, to enable **versatile and in-situ implementations**

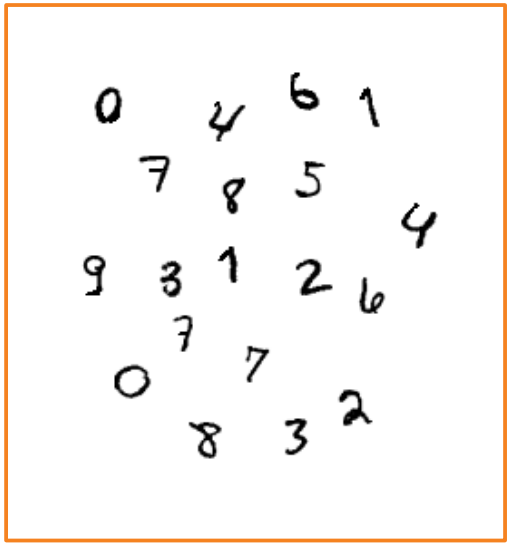
POWERFUL AUTOMATIC ANALYSIS WITH MACHINE LEARNING

Machine learning (1959): algorithms learn to carry out a task through experience

Example task: written digits classification



Training samples



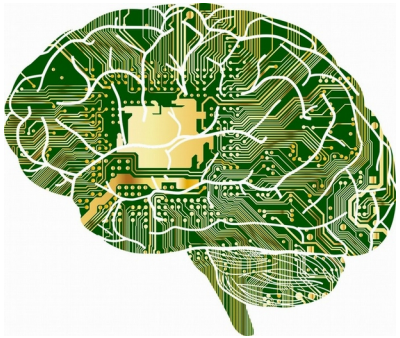
Test samples

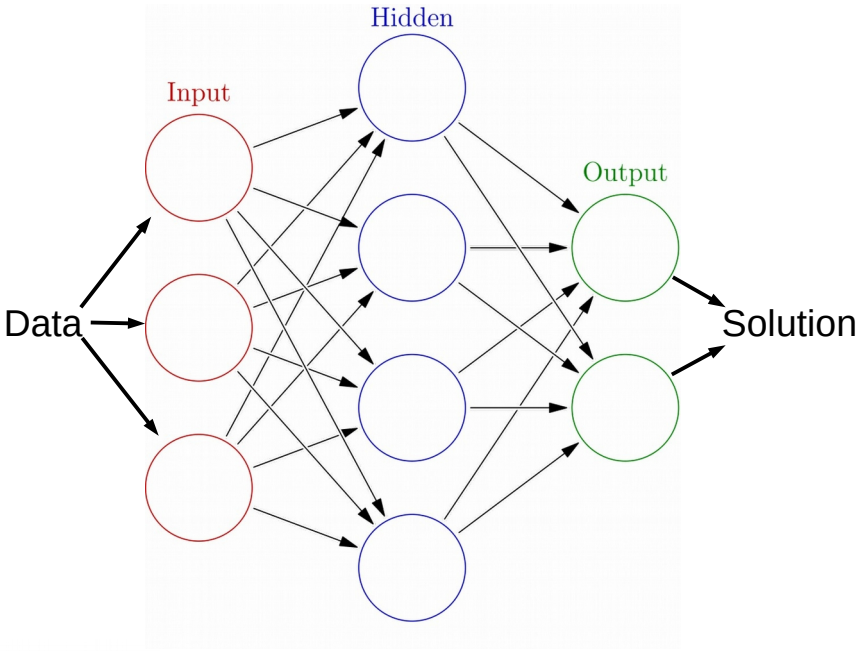
Training of machine learning model

Test of machine learning model

INSPIRED BY THE BRAIN

Neural network (NN) models have grown more and more powerful in the past decade, **outperforming humans** in complex tasks such as image and speech recognition, lip reading, chess, etc...

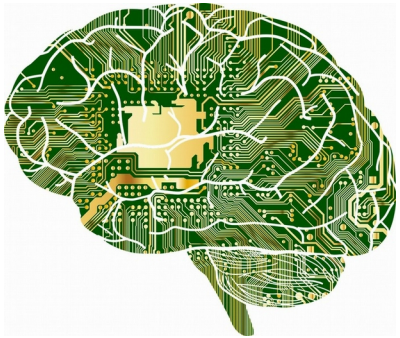


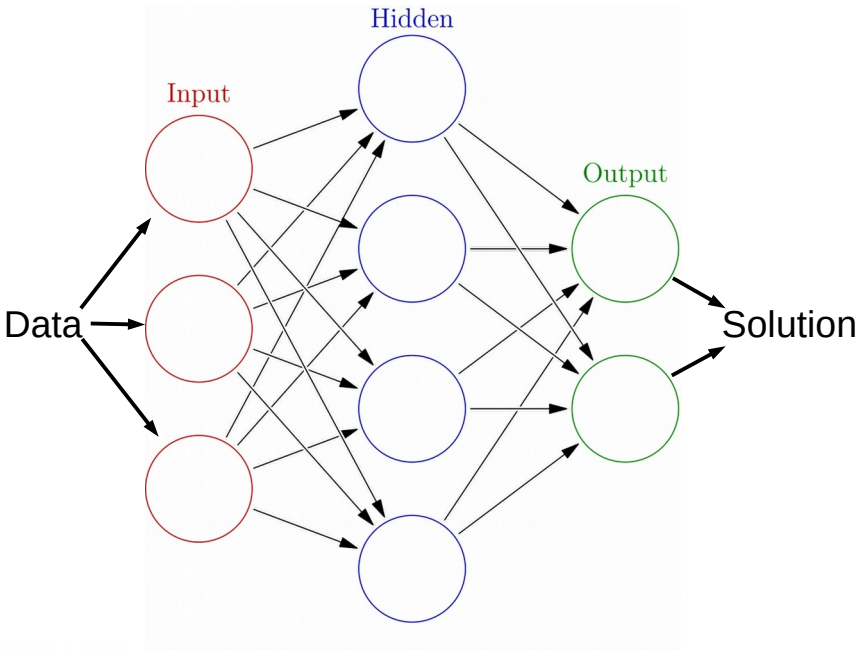


Images from Wikipedia.org

INSPIRED BY THE BRAIN

Neural network (NN) models have grown more and more powerful in the past decade, **outperforming humans** in complex tasks such as image and speech recognition, lip reading, chess, etc...





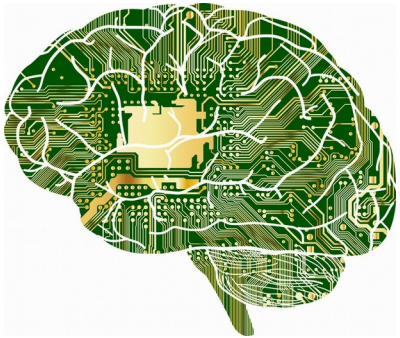
The larger the network, the higher the computational cost

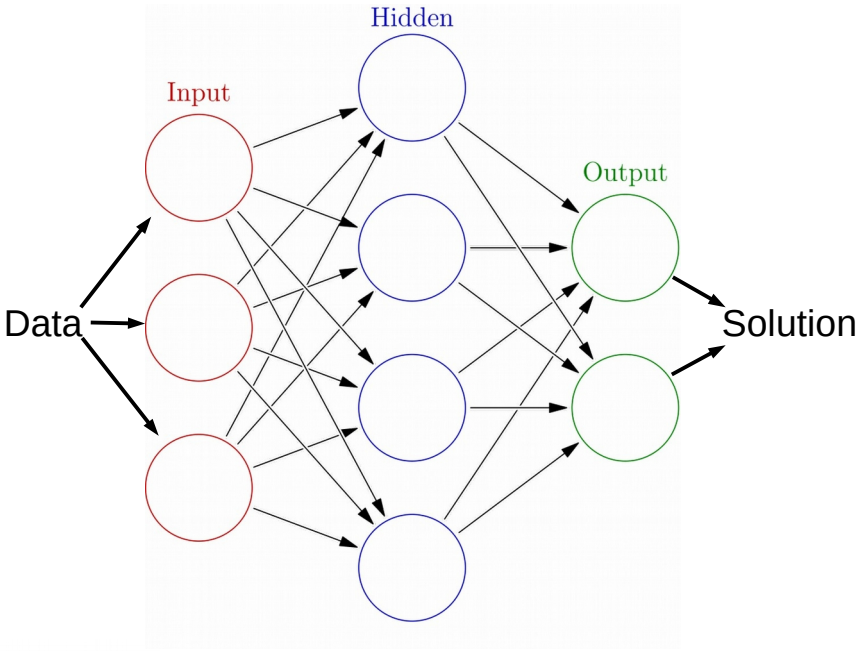
➔ trade-off between speed, compactness and cost

Images from Wikipedia.org

INSPIRED BY THE BRAIN

Neural network (NN) models have grown more and more powerful in the past decade, **outperforming humans** in complex tasks such as image and speech recognition, lip reading, chess, etc...





The larger the network, the higher the computational cost

➔ trade-off between speed, compactness and cost

Hardware-based NNs can greatly improve efficiency and speed

However they are usually difficult to train… we take a shortcut

Images from Wikipedia.org

CONVENTIONAL FLOW CYTOMETRY

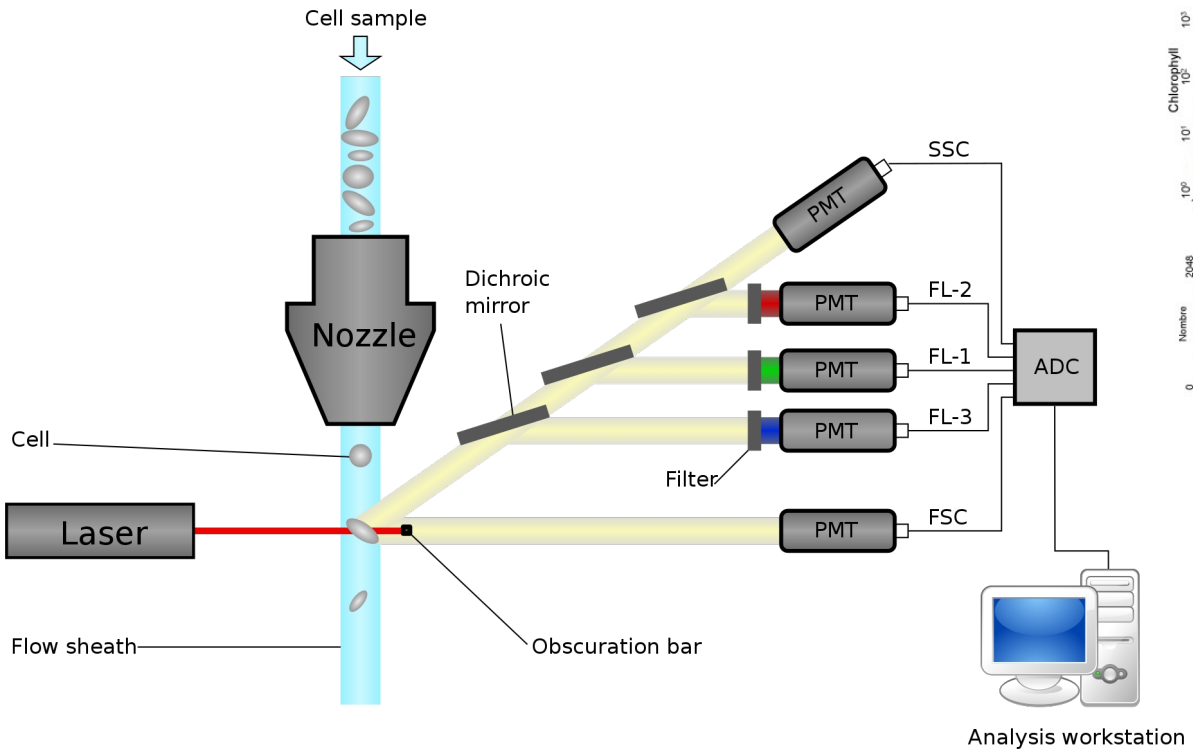


Image: https://commons.wikimedia.org/wiki/File:Cytometer.svg
License: https://creativecommons.org/licenses/by/3.0/deed.en

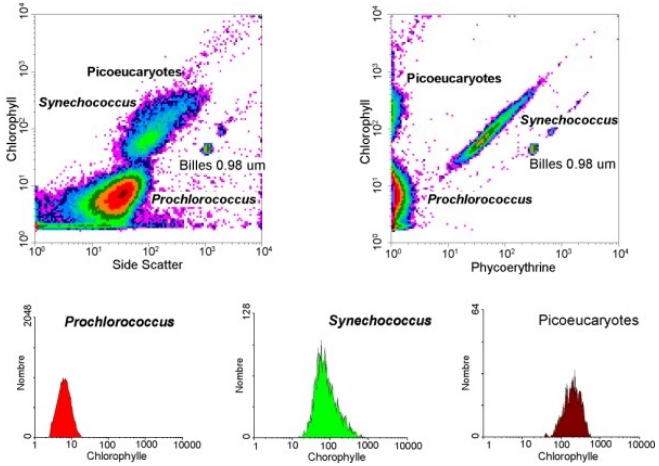


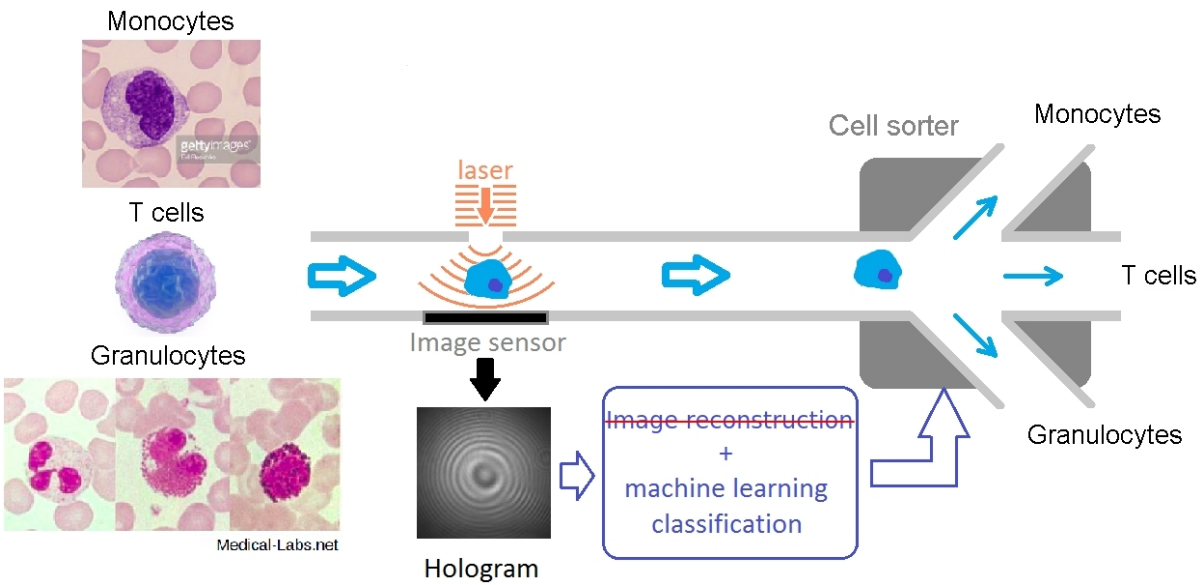
Image: https://en.wikipedia.org/wiki/Flow_cytometry#/media/File:Picoplancton_cytometrie.jpg
License: https://creativecommons.org/licenses/by-sa/2.5/deed.en

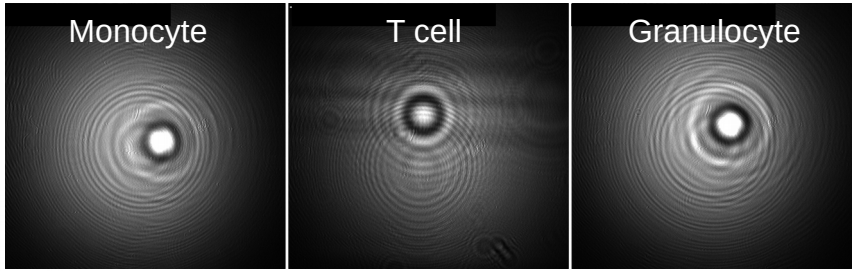
High throughput (~100,000 cell/s)

Fluorescent labels:

- often hinder live cell analysis
- additional cost and effort

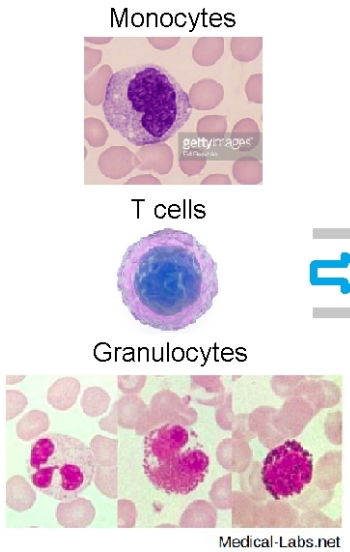
LABEL-FREE IMAGING FLOW CYTOMETRY

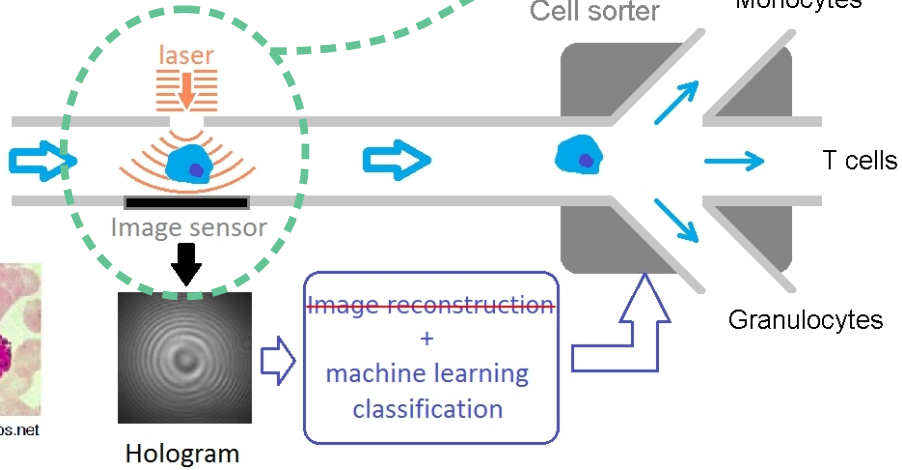


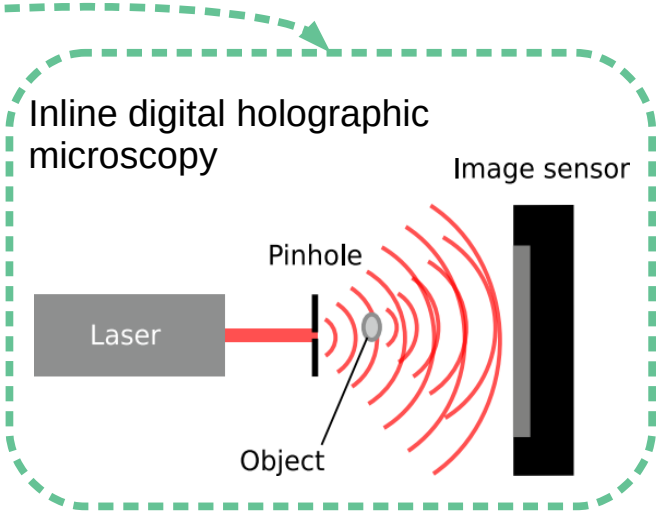


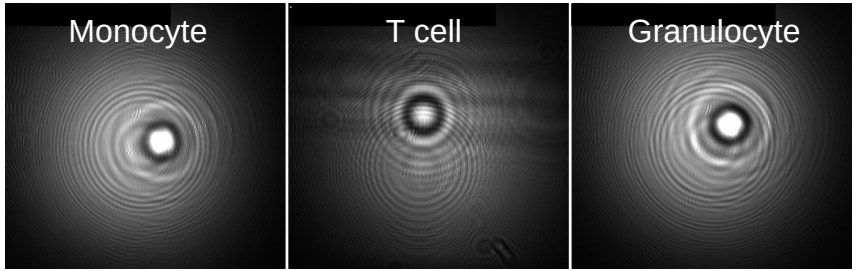
GHENT UNIVERSITY imec

LABEL-FREE IMAGING FLOW CYTOMETRY



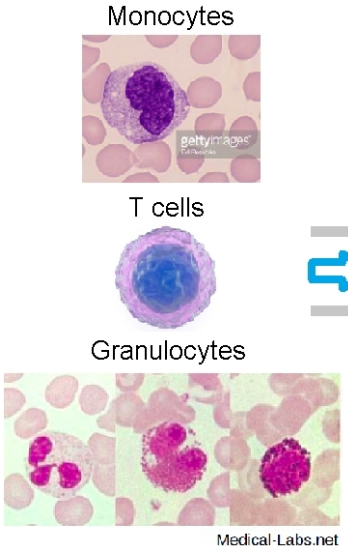


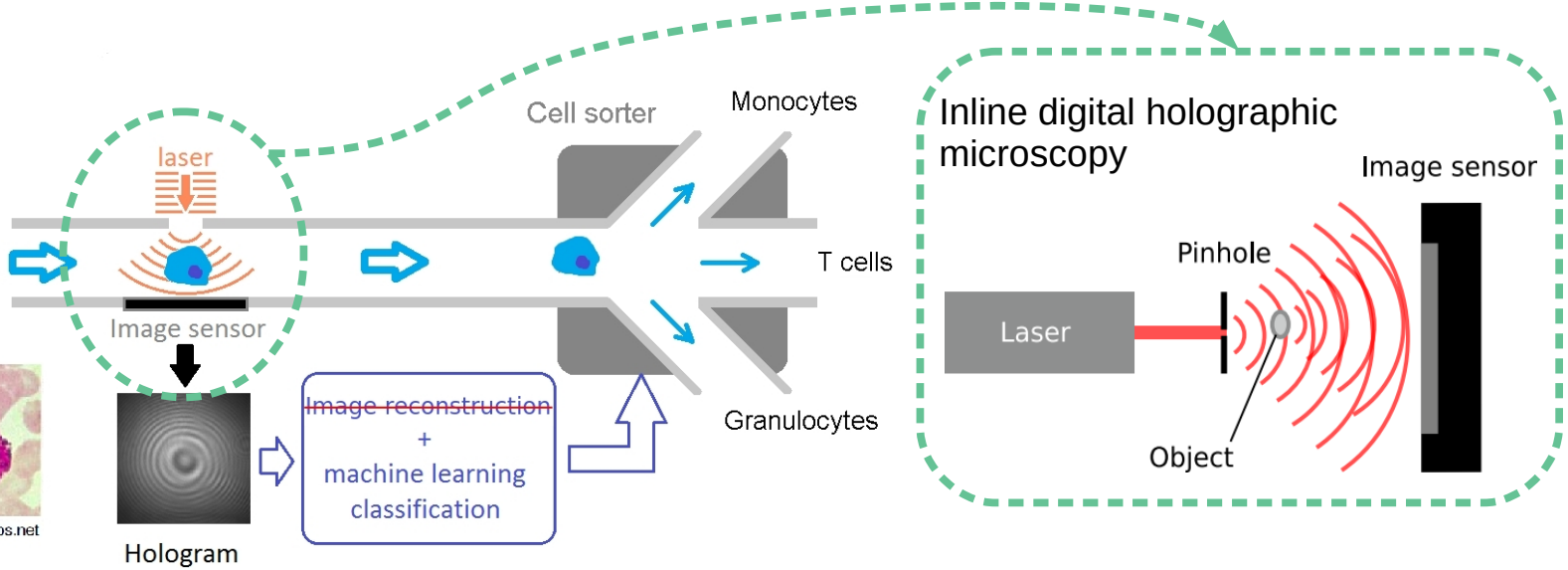


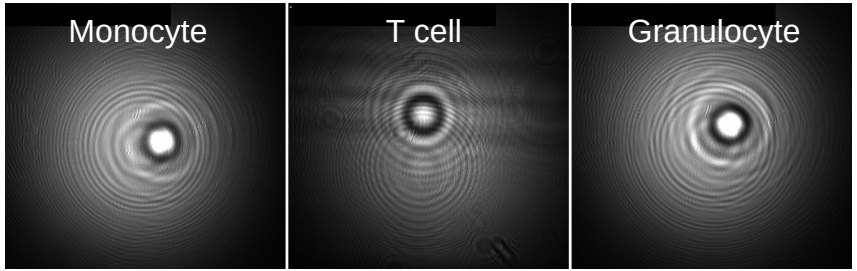


GHENT UNIVERSITY imec

LABEL-FREE IMAGING FLOW CYTOMETRY





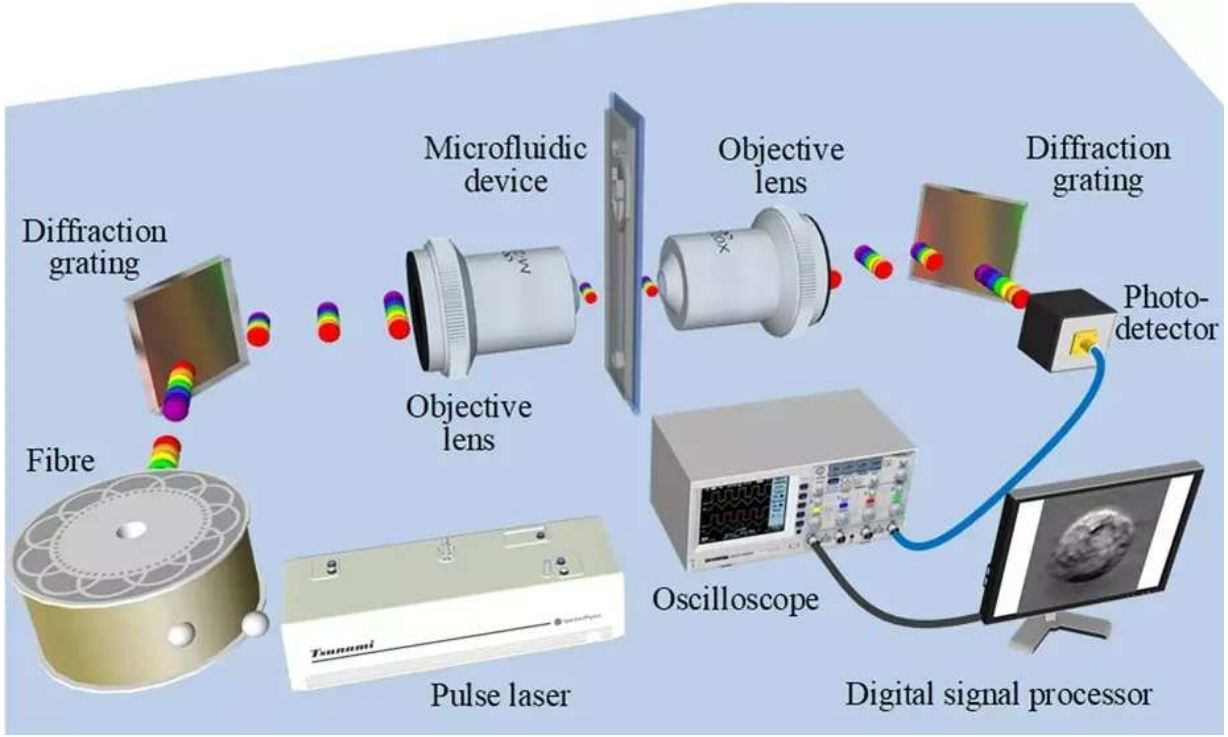


Main bottlenecks:

- **computational cost** of classification
- camera frame rate (up to ~1000 particles/s)
  - ➔ but several holograms could be acquired in parallel

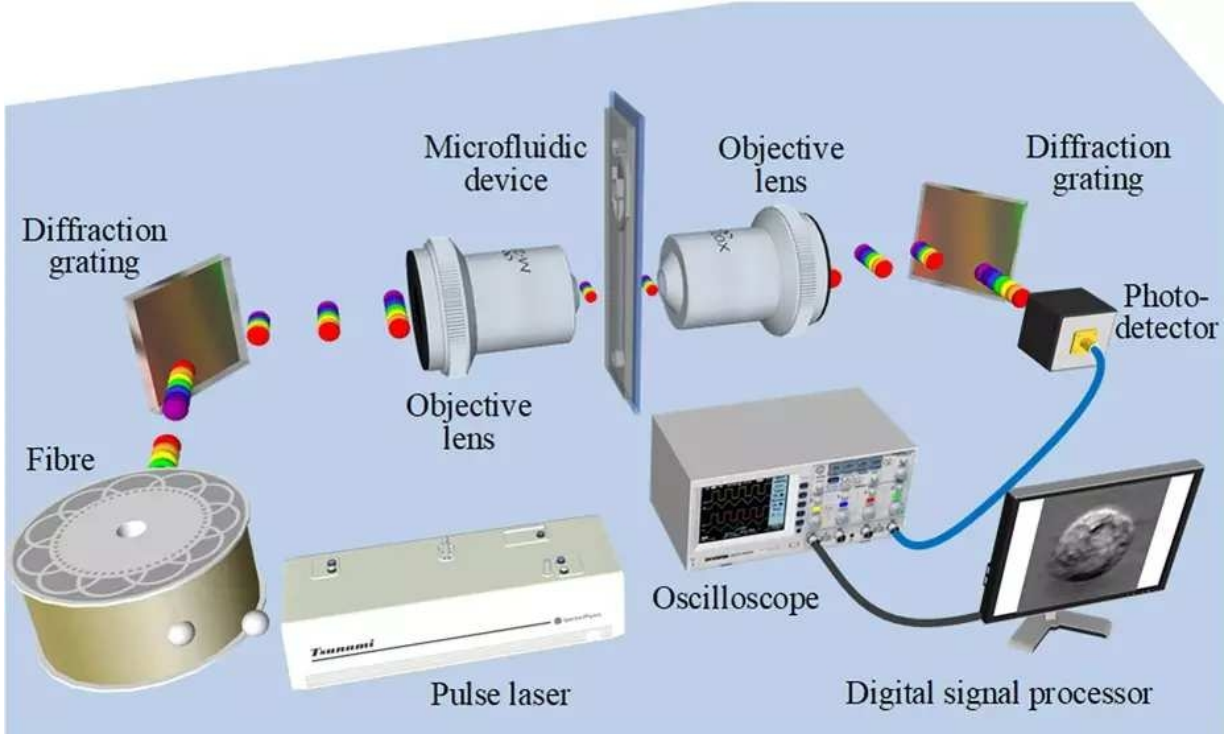
GHENT UNIVERSITY imec

HIGH THROUGHPUT IMAGING WITH OPTOFLUIDIC TIME-STRETCH MICROSCOPY



Hirofumi Kobayashi, et al. “Label-free detection of cellular drug responses by high-throughput bright-field imaging and machine learning”. *Scientific reports*, 2017

HIGH THROUGHPUT IMAGING WITH OPTOFLUIDIC TIME-STRETCH MICROSCOPY



Hirofumi Kobayashi, et al. “Label-free detection of cellular drug responses by high-throughput bright-field imaging and machine learning”. *Scientific reports*, 2017

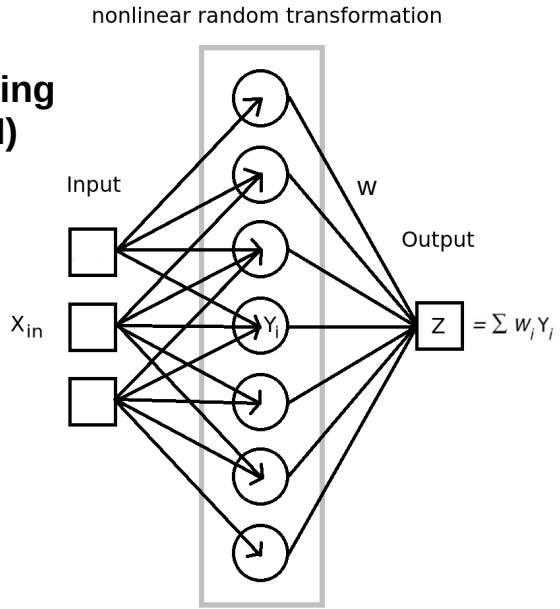
- Very high-throughput: up to 100,000 cells/s

But...

- Relatively expensive and complicated
- ~1Tbit/s of continuous measurement data!
  - online operation is desiderable!
  - necessary for cell sorting
  - **need for computationally cheap analysis**

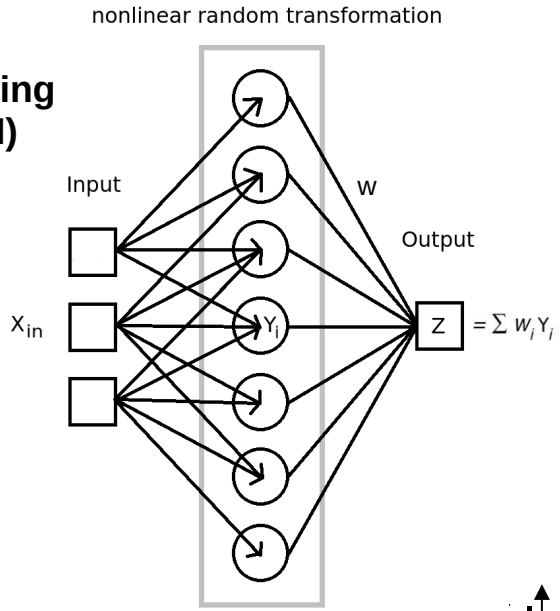
A SHORTCUT TO EXPLOIT HARDWARE ‘COMPUTATION’

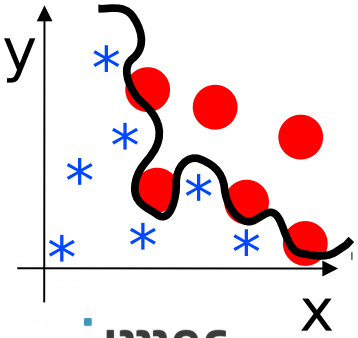
Extreme learning machine (ELM)



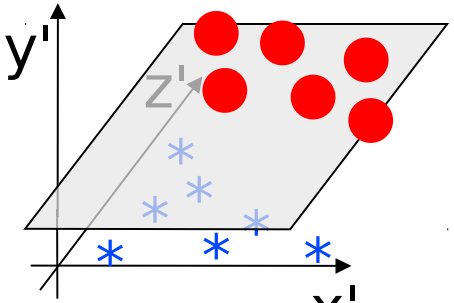
HARDWARE-BASED RANDOM DIMENSIONALITY EXPANSION

Extreme learning machine (ELM)





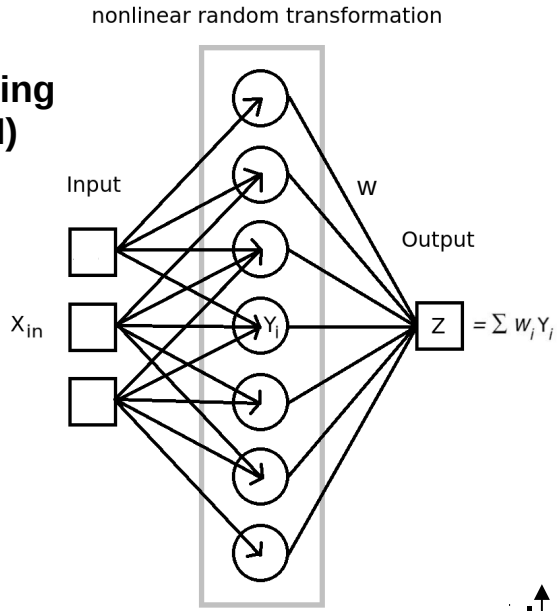
Map to a higher dimensional space

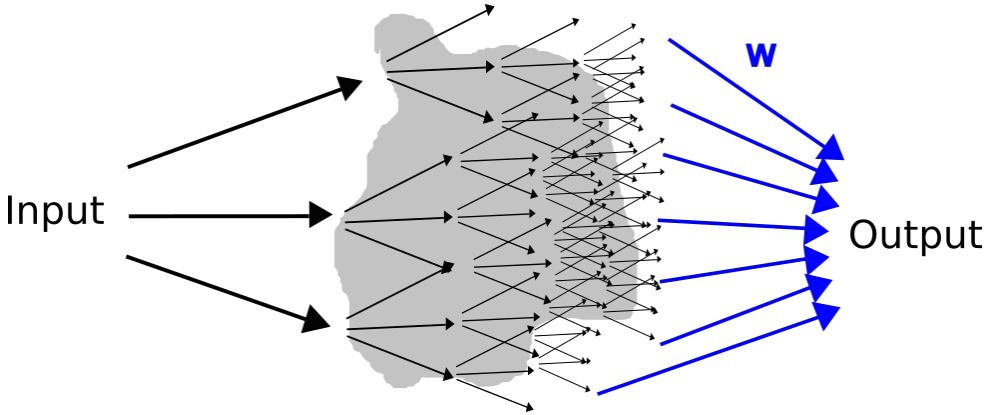


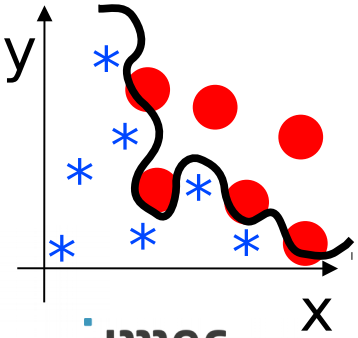
Linearly separable

HARDWARE-BASED RANDOM DIMENSIONALITY EXPANSION

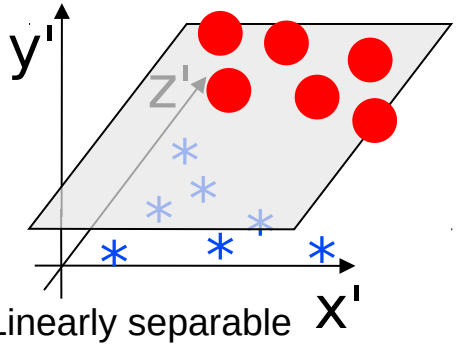
Extreme learning machine (ELM)







Map to a higher dimensional space



Suitable for **hardware implementations**

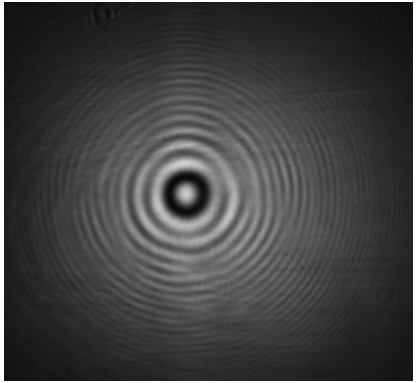
Photonics: high-speed, efficient parallel processing

GHENT UNIVERSITY imec

WHITE BLOOD CELL HOLOGRAM CLASSIFICATION

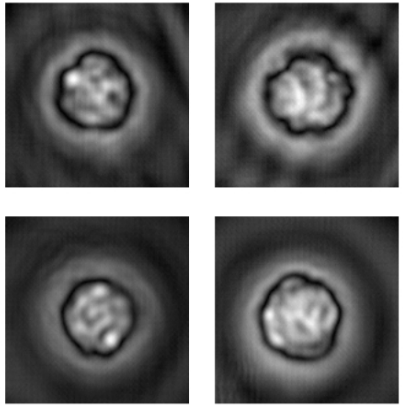
REAL CELL HOLOGRAMS

Raw hologram

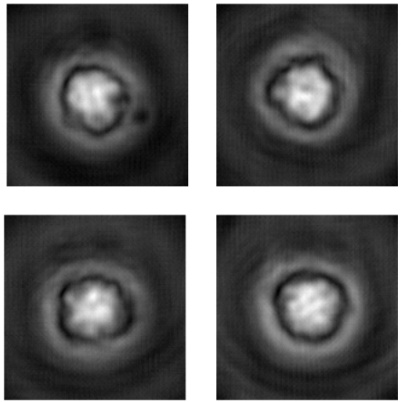


Reconstruction

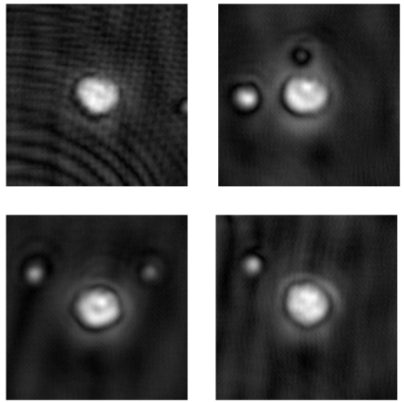
Granulocytes



Monocytes



T-lymphocytes



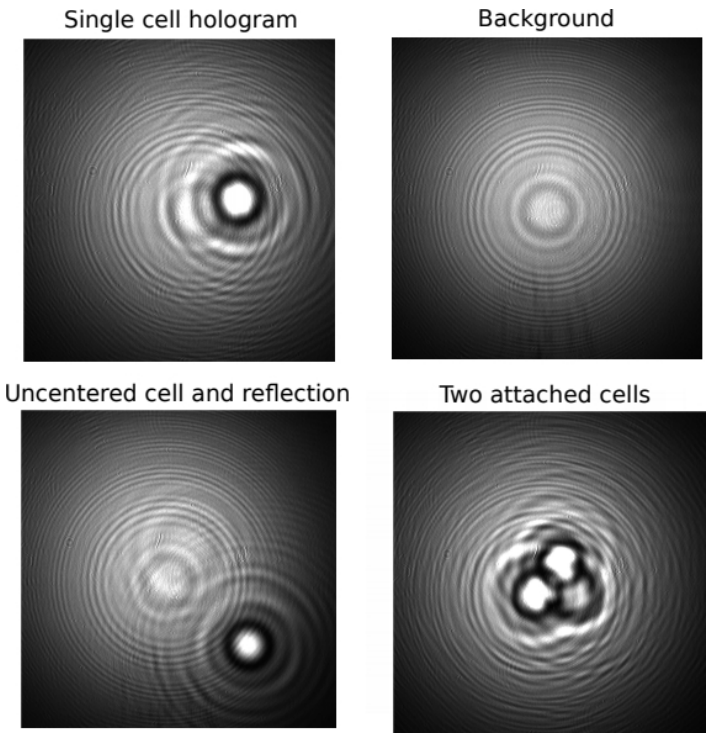
WBC holograms from Imec collaborators:

- 20,797 monocyte
- 3,753 T cell
- 32,514 granulocyte

Goal: fast classification

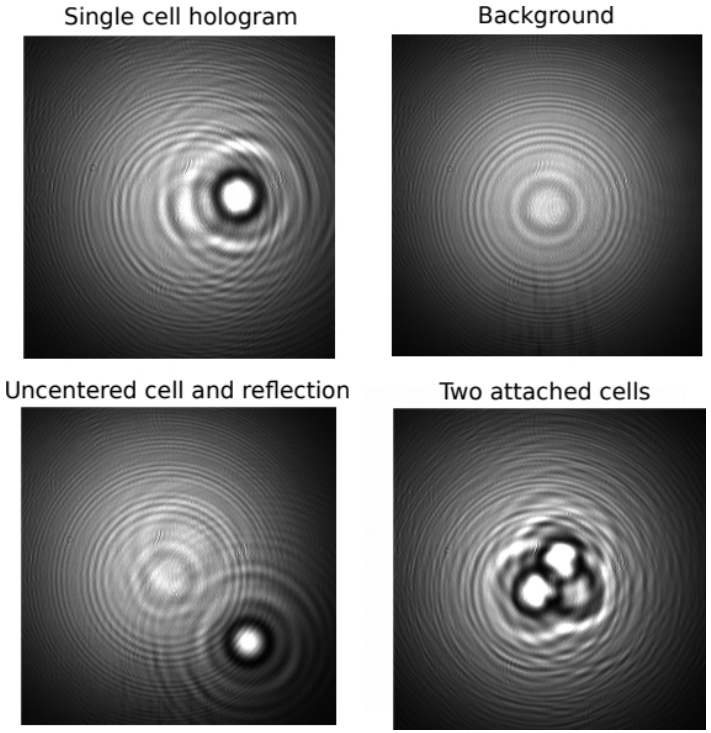
"Fast and robust Fourier domain-based classification for on-chip lens-free flow cytometry," Bruno Cornelis et al, *Optics Express* (2018)

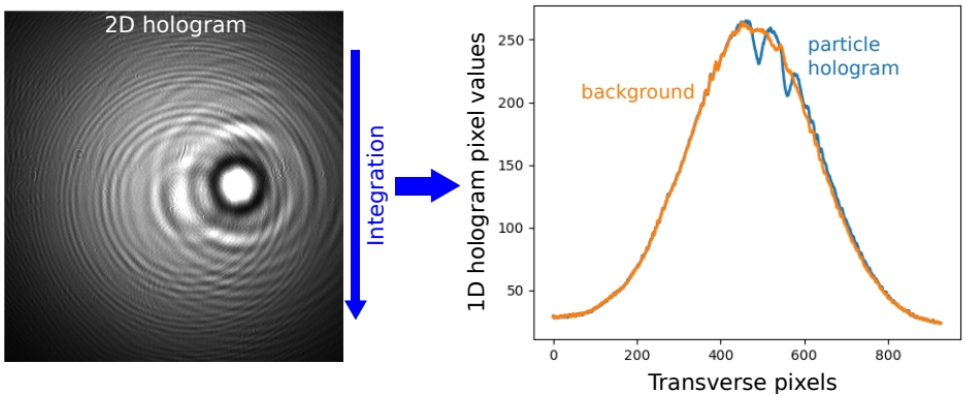
ADDRESSING HOLOGRAM VARIABILITY (NOISE)



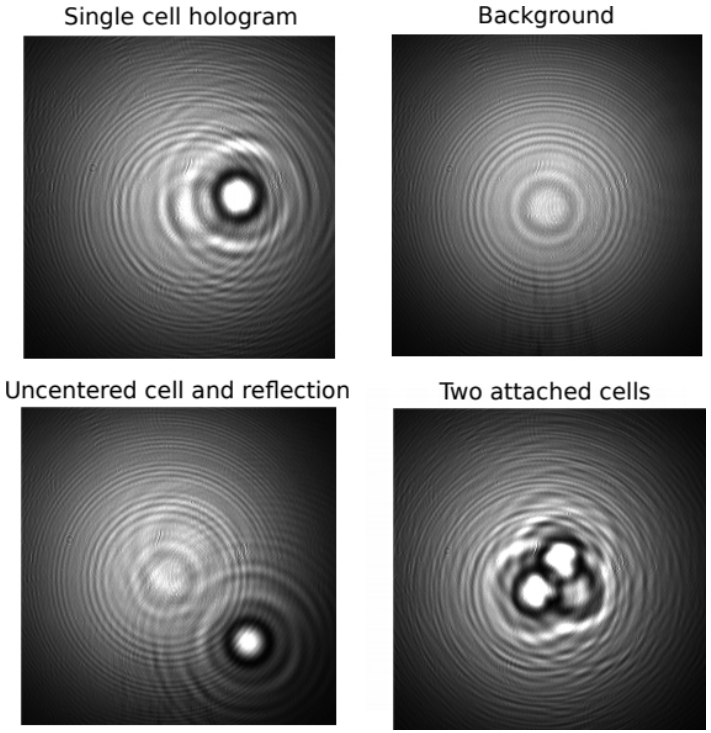
GHENT UNIVERSITY imec

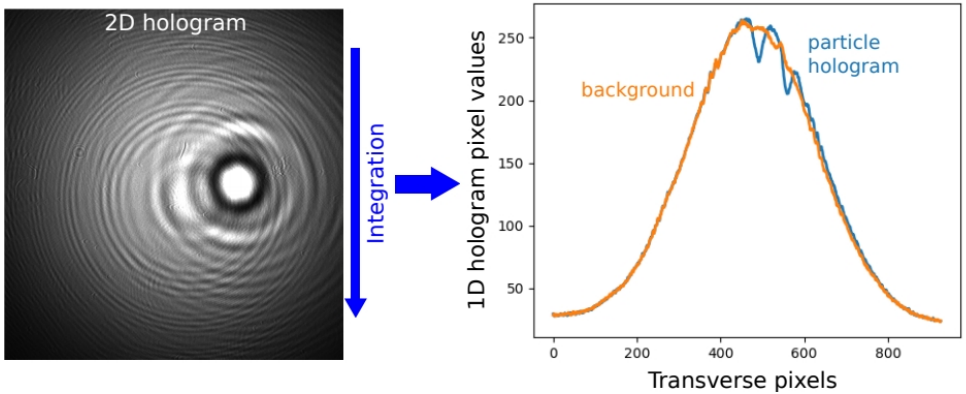
ADDRESSING HOLOGRAM VARIABILITY (NOISE)

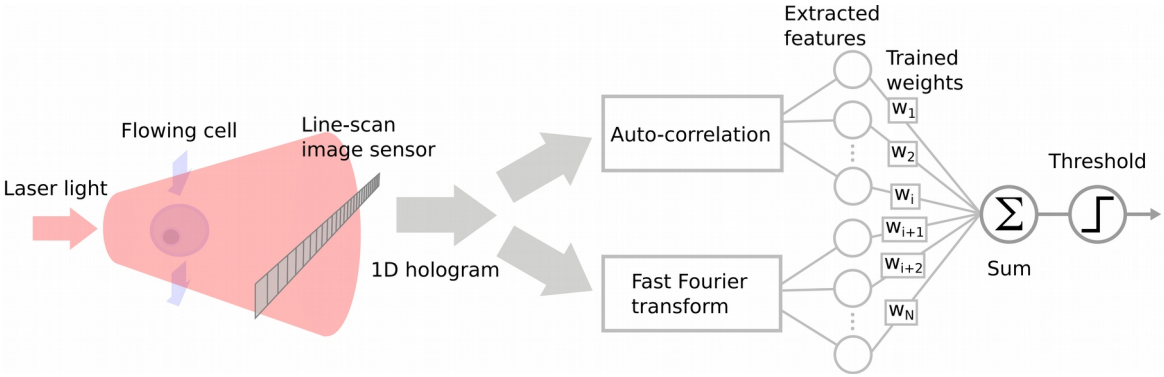




ADDRESSING HOLOGRAM VARIABILITY (NOISE)

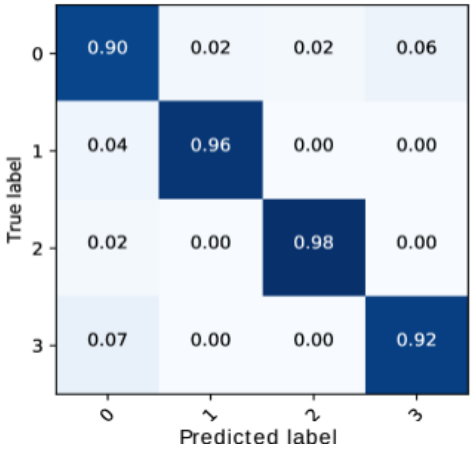






SHORTCUT LEARNING!

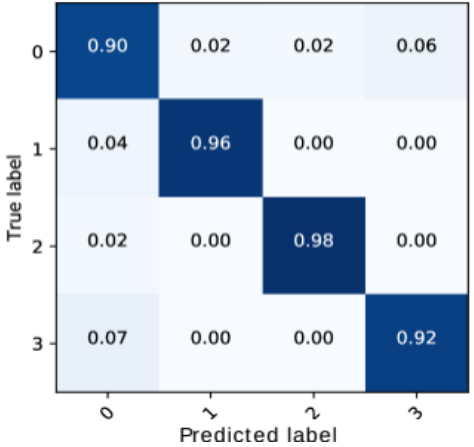
WBC classification



High accuracy!

SHORTCUT LEARNING!

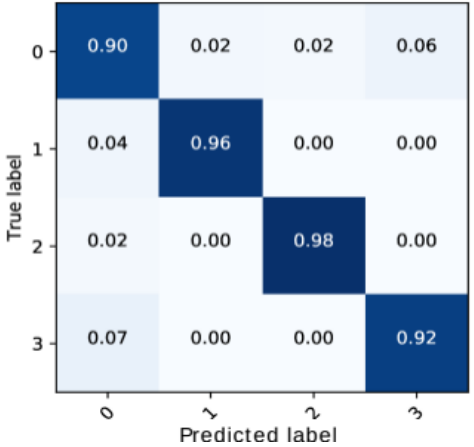
WBC classification



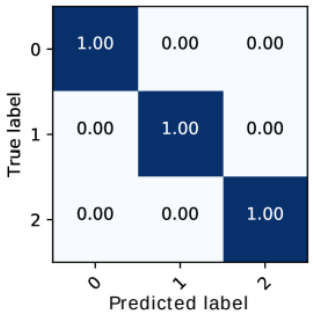
High accuracy!

SHORTCUT LEARNING!

WBC classification



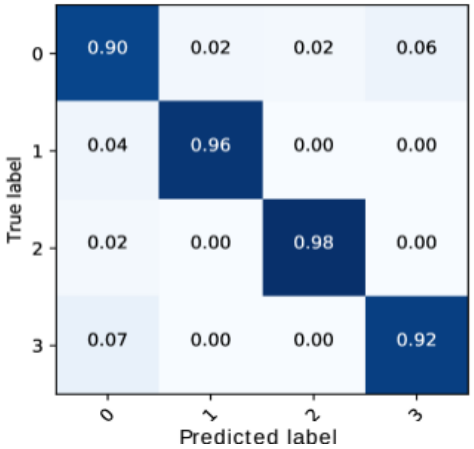
Only backgrounds



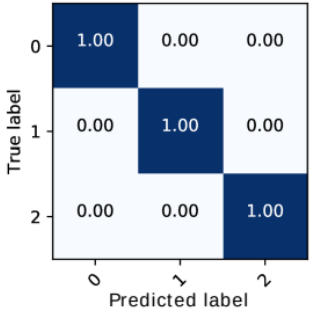
High accuracy!

SHORTCUT LEARNING!

WBC classification



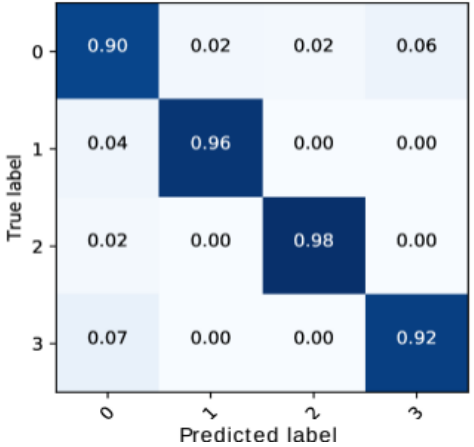
Only backgrounds



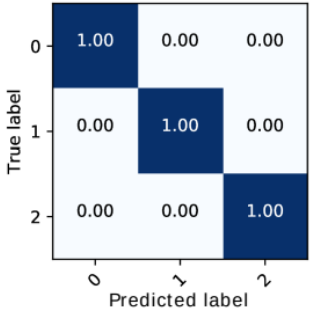
High accuracy!

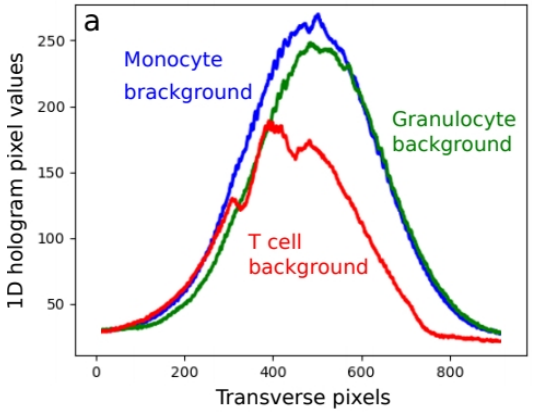
SHORTCUT LEARNING!

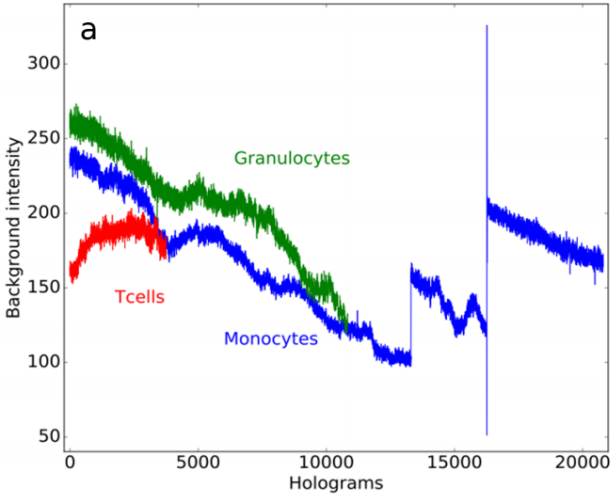
WBC classification



Only backgrounds

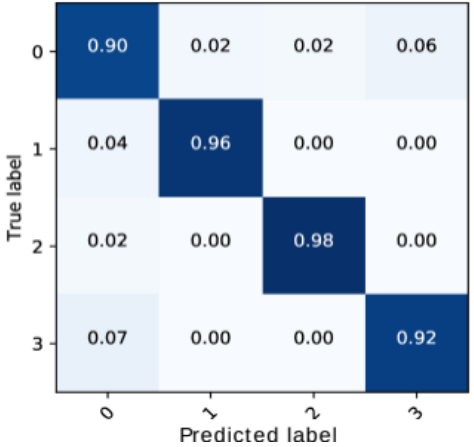




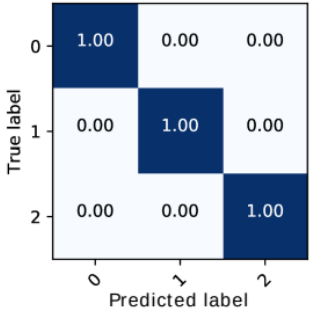


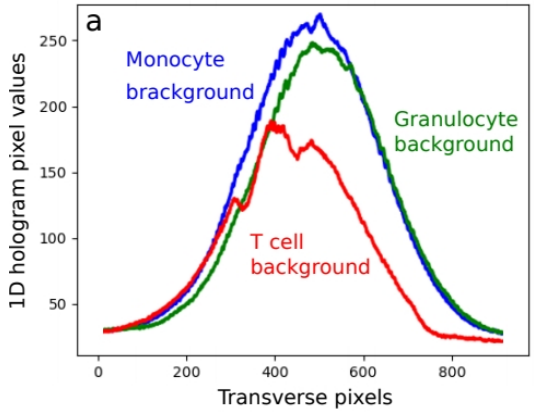
SHORTCUT LEARNING!

WBC classification



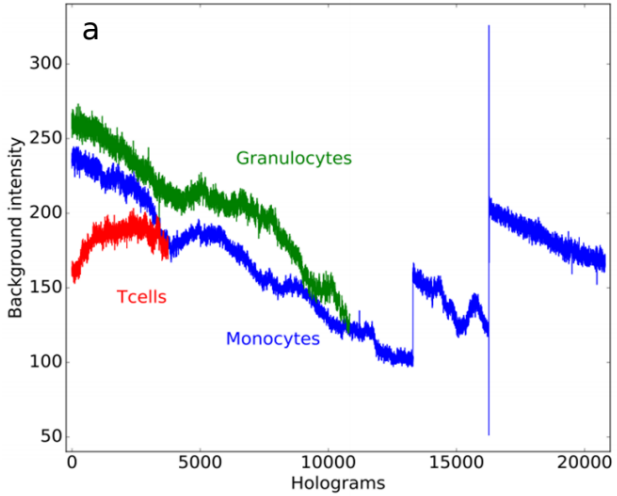
Only backgrounds





Shortcut learning (*measurement bias*): often ignored or underestimated

Geirhos, R., et al. "Shortcut learning in deep neural networks." *Nat. Mach. Intell.*, 2020



- Cross-validation does not help
- Background subtraction is not sufficient

$$h_i - h_{i-1} = \delta h_i^{\text{bkgr}} + \mathcal{N}(h_i^{\text{bkgr}}, C_i)$$

GHENT UNIVERSITY imec

TREATMENT OF MEASUREMENT BIAS

Measurement bias is a two-fold problem:

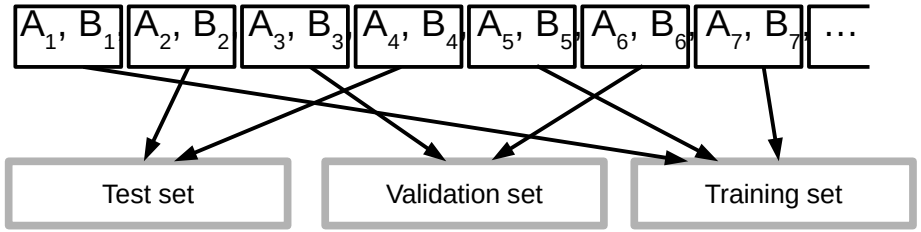
1) undermines learning
2) test results are inflated

TREATMENT OF MEASUREMENT BIAS

Measurement bias is a two-fold problem:

1) undermines learning
2) test results are inflated

Intertwined class measurements address both:
$A_1, B_1, A_2, B_2, A_3, B_3, A_4, B_4, A_5, B_5, A_6, B_6, A_7, B_7, \ldots$

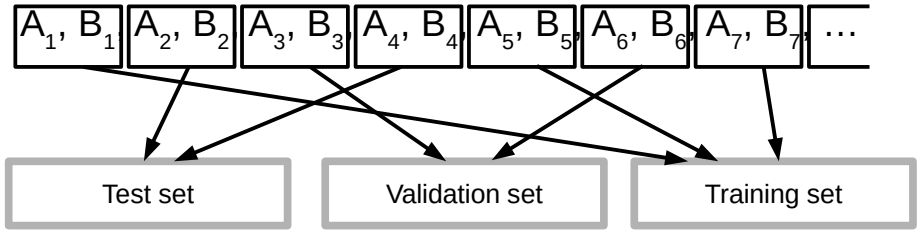


TREATMENT OF MEASUREMENT BIAS

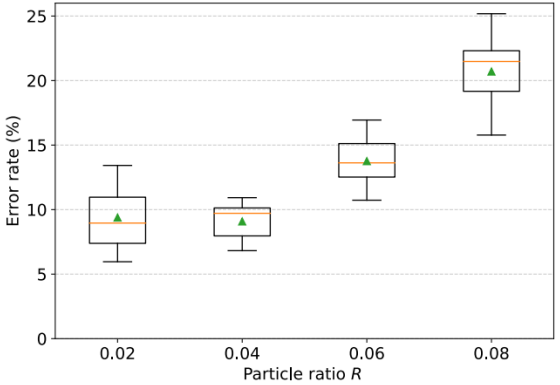
Measurement bias is a two-fold problem:

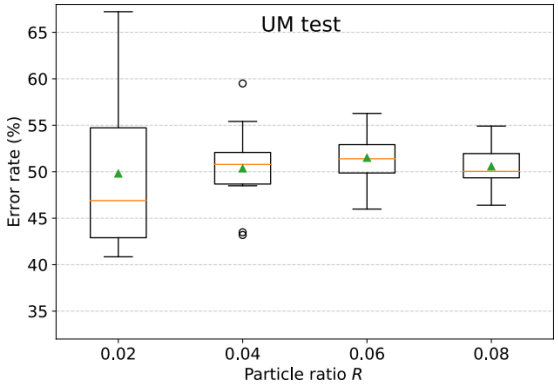
1) undermines learning
2) test results are inflated

Intertwined class measurements address both:
$A_1, B_1, A_2, B_2, A_3, B_3, A_4, B_4, A_5, B_5, A_6, B_6, A_7, B_7, \ldots$



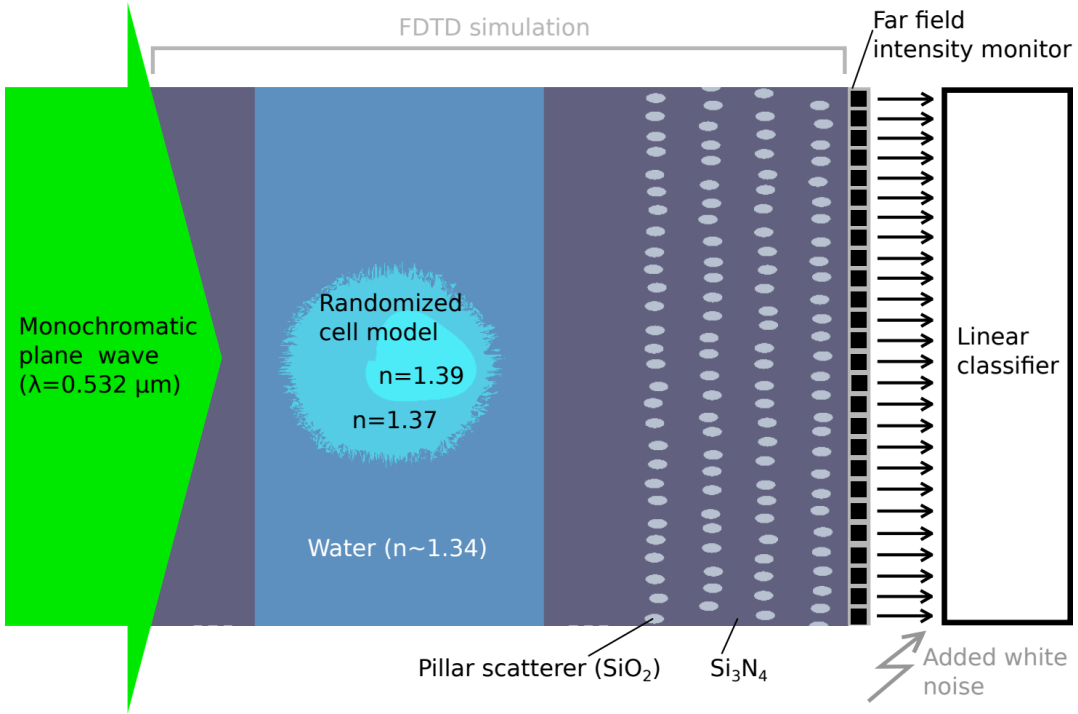
Effectiveness demonstrated in dedicated experiment with microspheres:





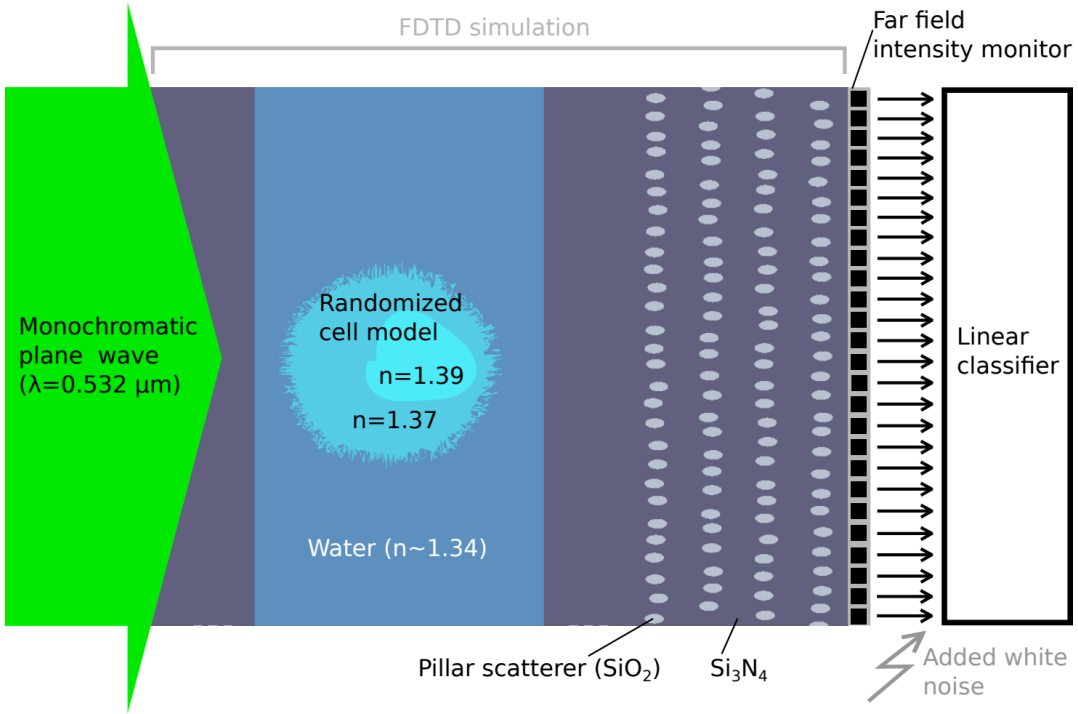
DIMENSIONALITY EXPANSION WITH DIELECTRIC SCATTERERS

RANDOMIZED SIMULATIONS OF CELL ILLUMINATION

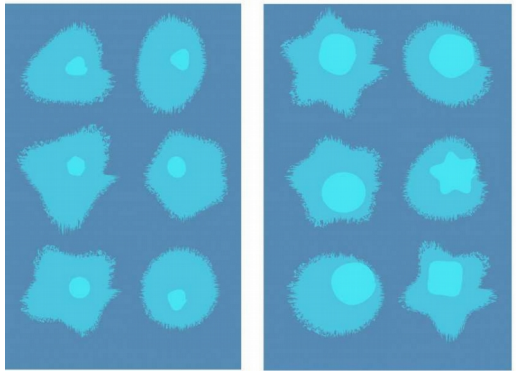


7200 FDTD simulations per scatterer configuration

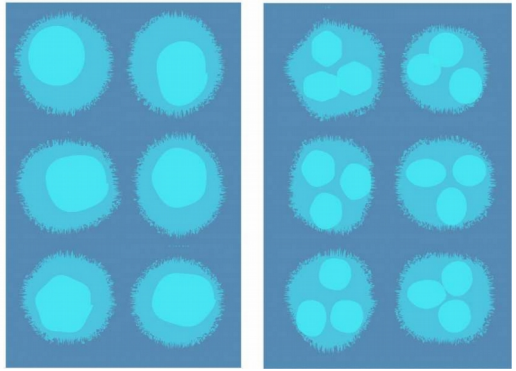
RANDOMIZED SIMULATIONS OF CELL ILLUMINATION



Task 1: nucleus size



Task 2: nucleus shape

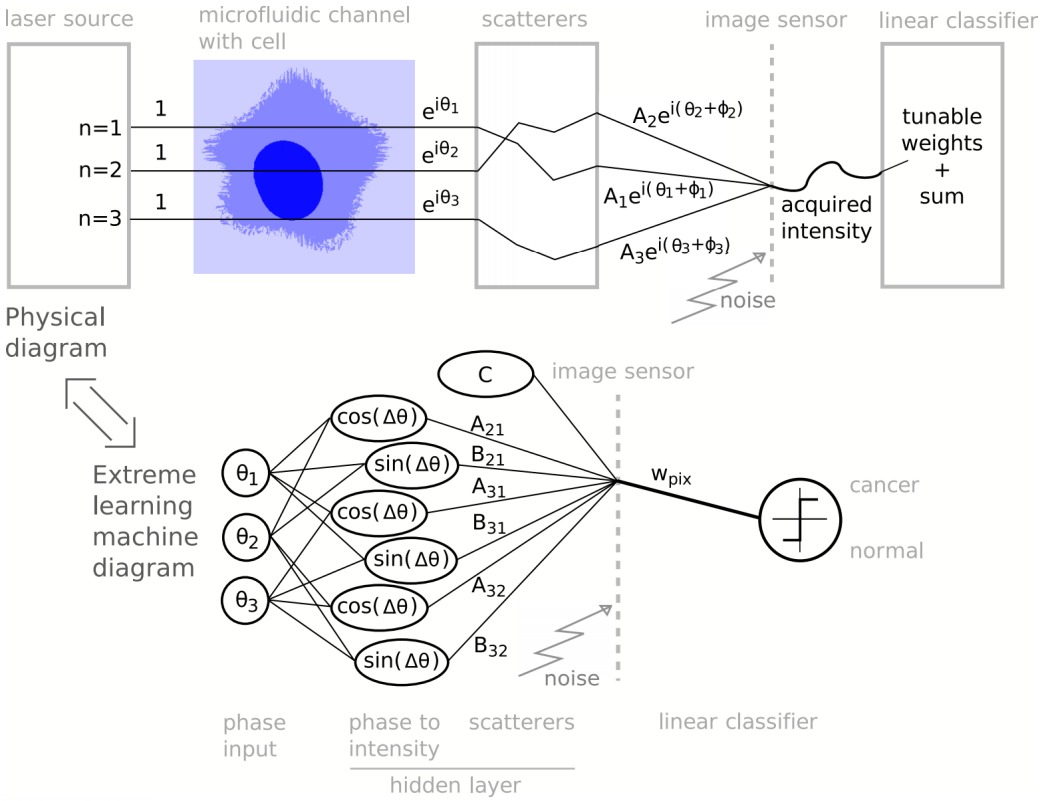


7200 FDTD simulations per scatterer configuration

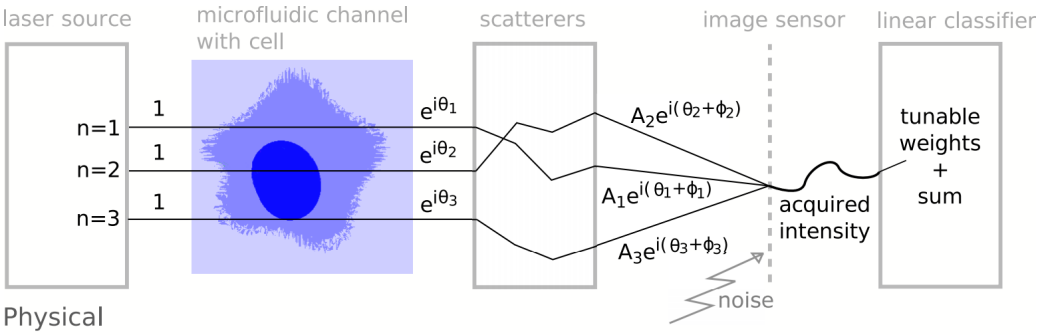
GHENT UNIVERSITY

imec

NONLINEARITY AND ELM EQUIVALENCE

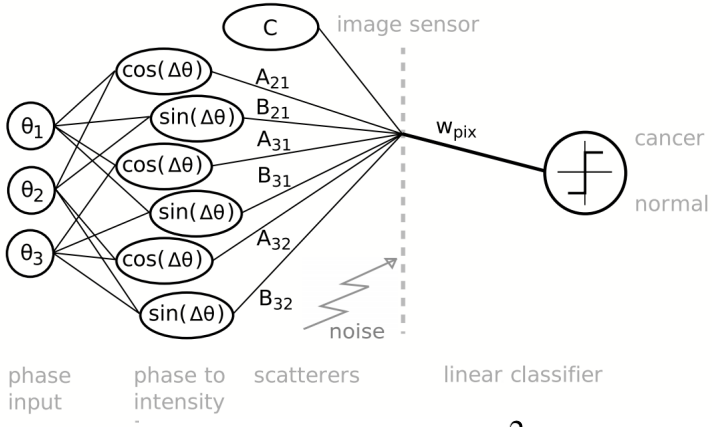


NONLINEARITY AND ELM EQUIVALENCE



Physical diagram

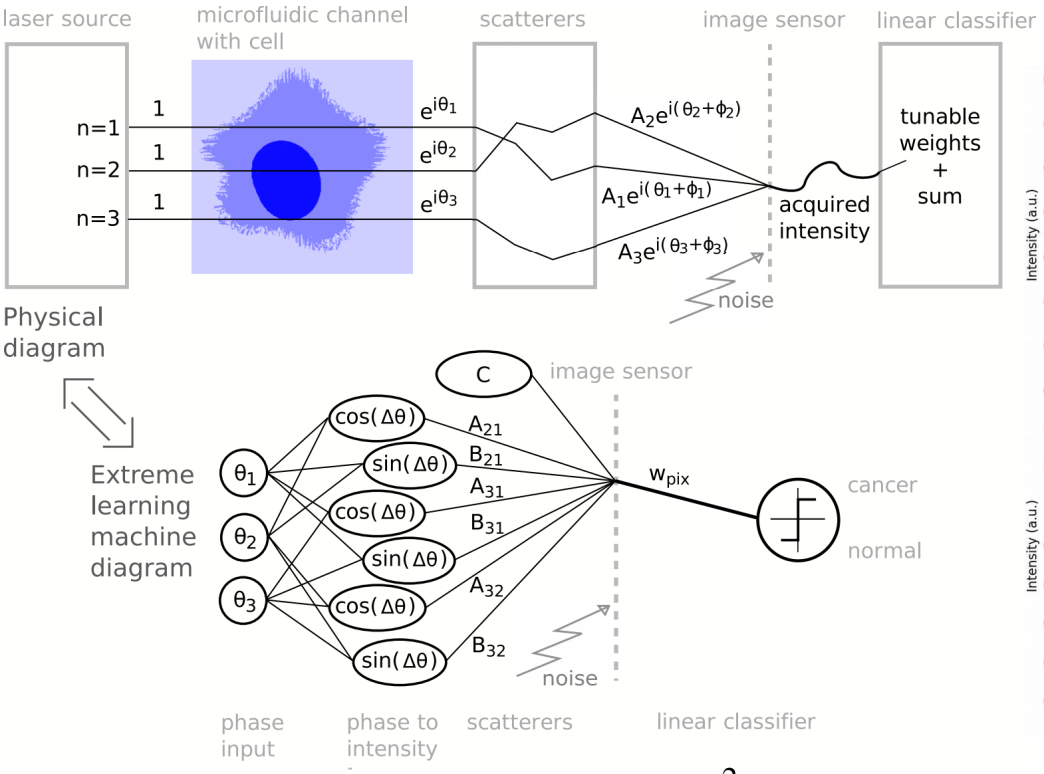
Extreme learning machine diagram

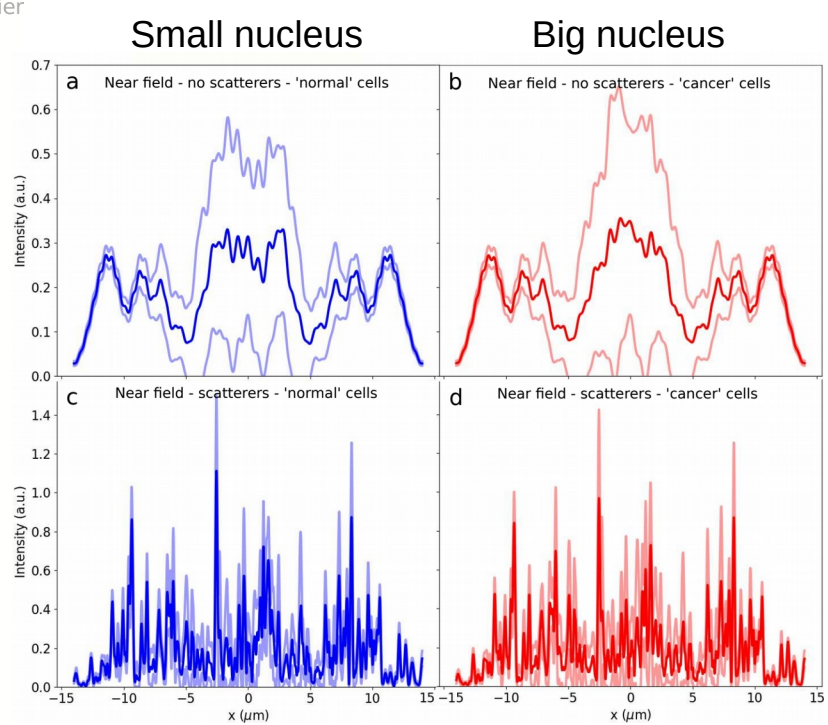


$$I \propto \left| \sum_n A_n e^{i(\theta_n+\phi_n)} \right|^2 = C + \sum_{m<n} \left[A_{nm}\cos(\theta_n-\theta_m) + B_{nm}\sin(\theta_n-\theta_m) \right]$$

GHENT UNIVERSITY imec

NONLINEARITY AND ELM EQUIVALENCE

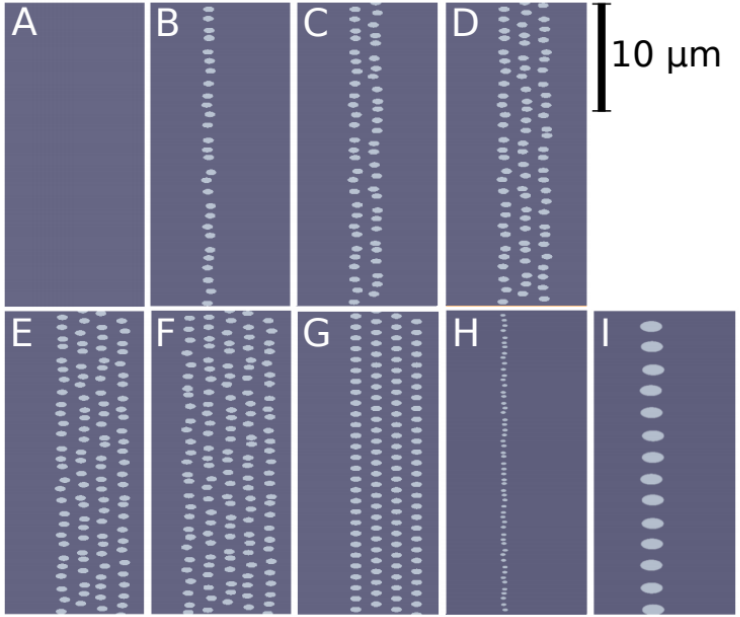


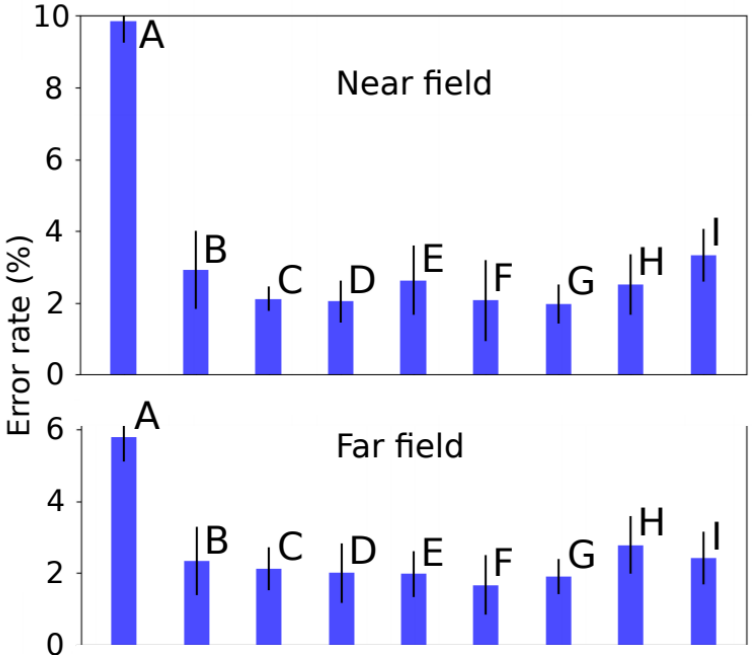


$$I \propto \left| \sum_n A_n e^{i(\theta_n+\phi_n)} \right|^2 = C + \sum_{m<n} \left[A_{nm}\cos(\theta_n-\theta_m) + B_{nm}\sin(\theta_n-\theta_m) \right]$$

GHENT UNIVERSITY imec

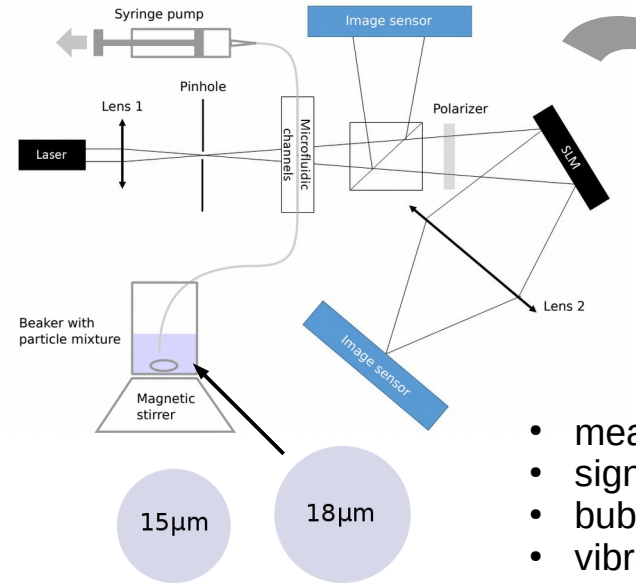
EXPLORATION OF SCATTERER CONFIGURATION

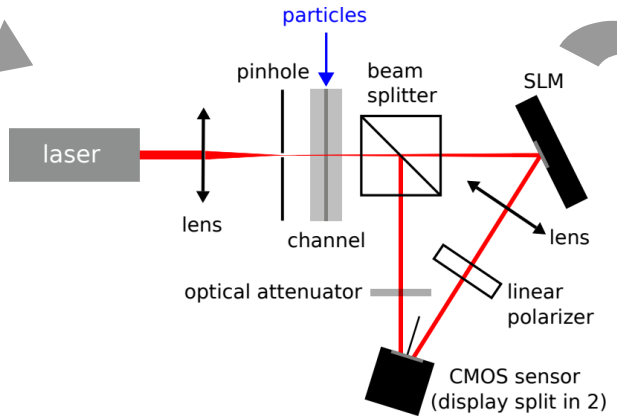


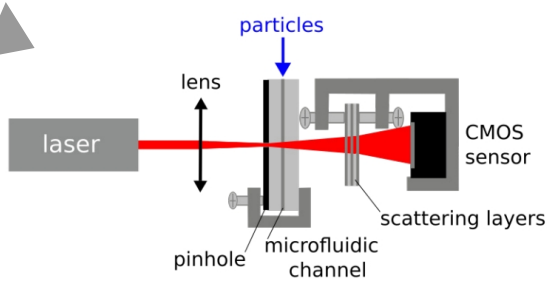


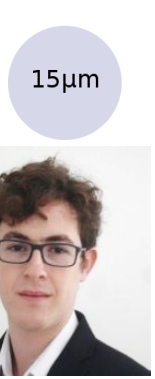
DEVELOPMENT OF FLOW CYTOMETRY EXPERIMENT

SETUP EVOLUTION





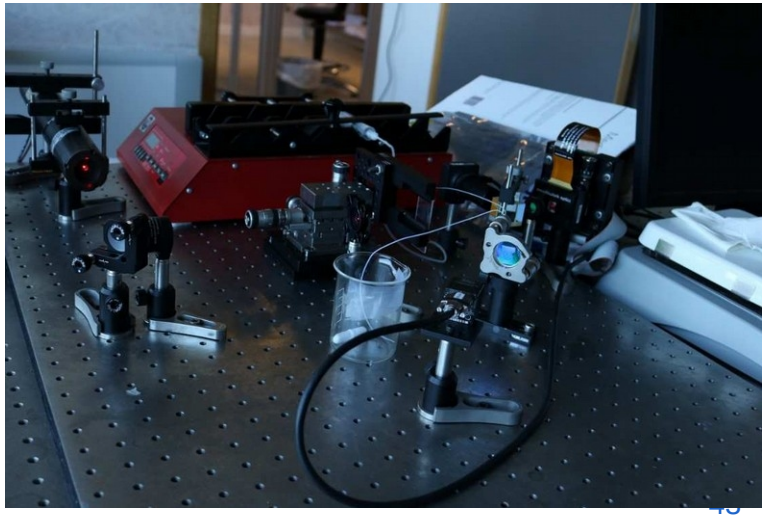




Jeremy Vatin
Centrale Supélec

Emmanuel Gooskens
UGent

- measurement bias
- signal-to-background ratio
- bubbles or dirt
- vibrations
- background detection and subtraction
- SLM flicker
- feature selection
- exploration and selection of scattering media
- few-samples estimation
- motion blur

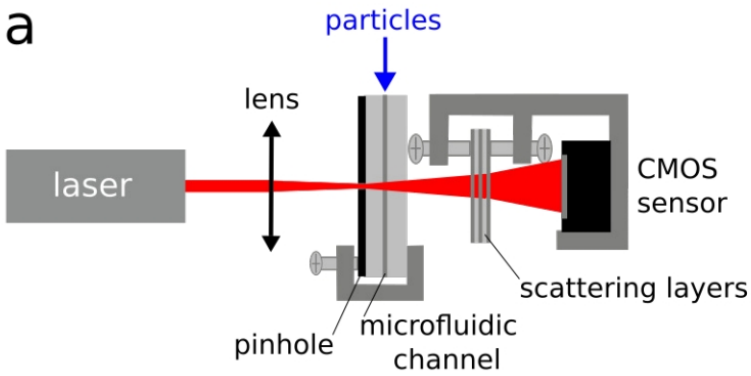


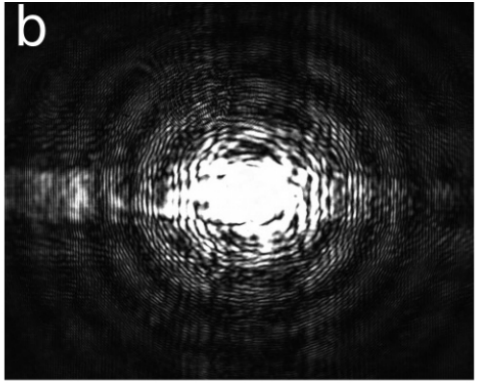
GHENT UNIVERSITY imec

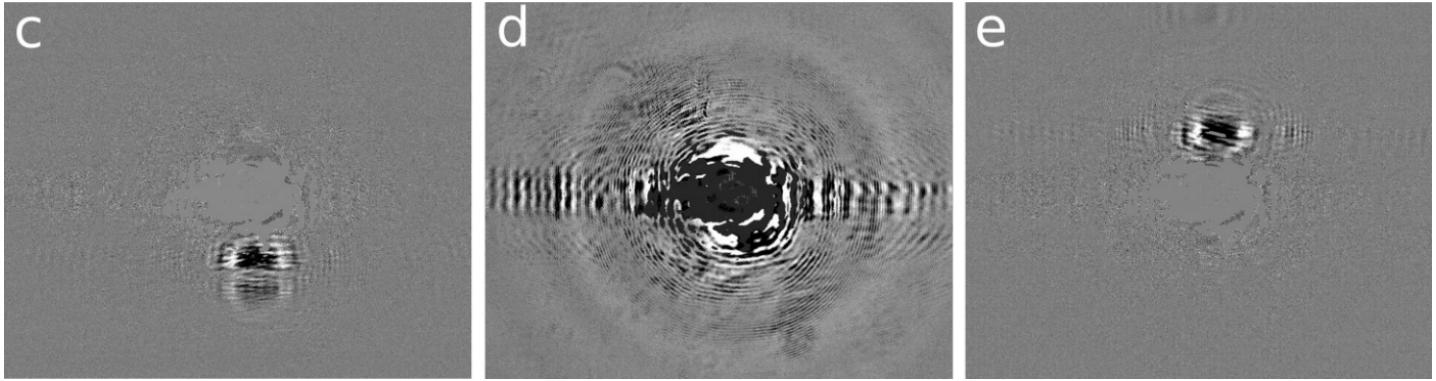
HIGH SNR IN ACQUIRED PATTERNS

15µm

18µm

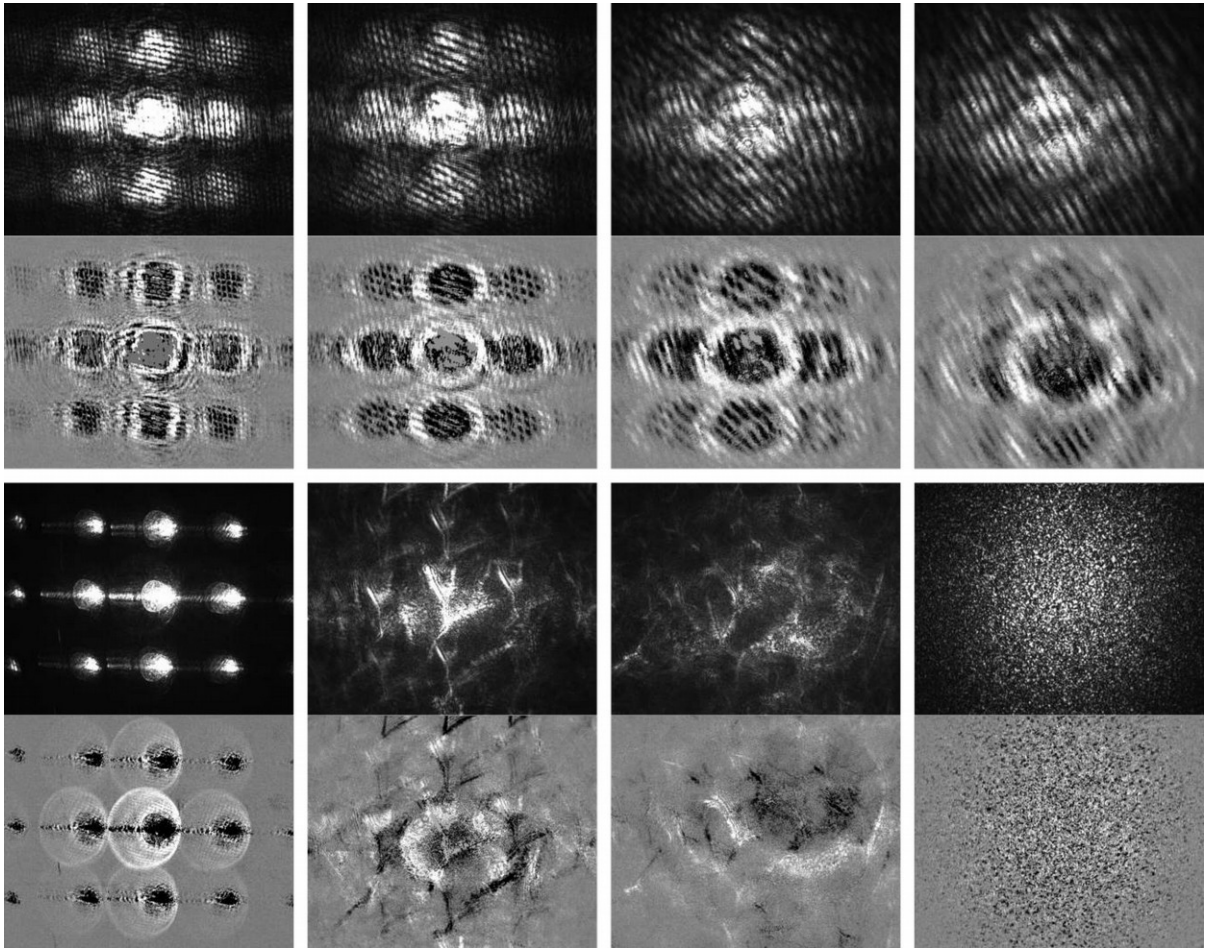






EXPLORATION OF SCATTERING LAYERS

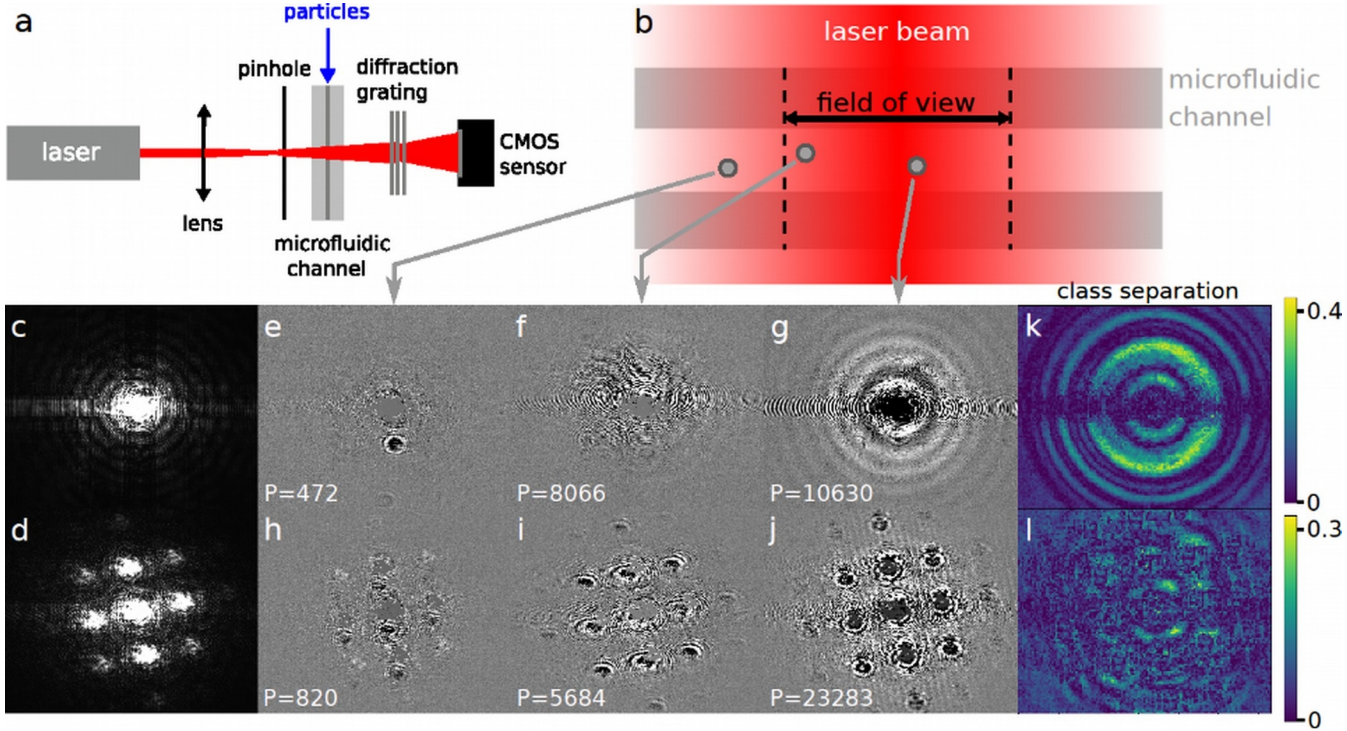
More than 40 configurations tested!



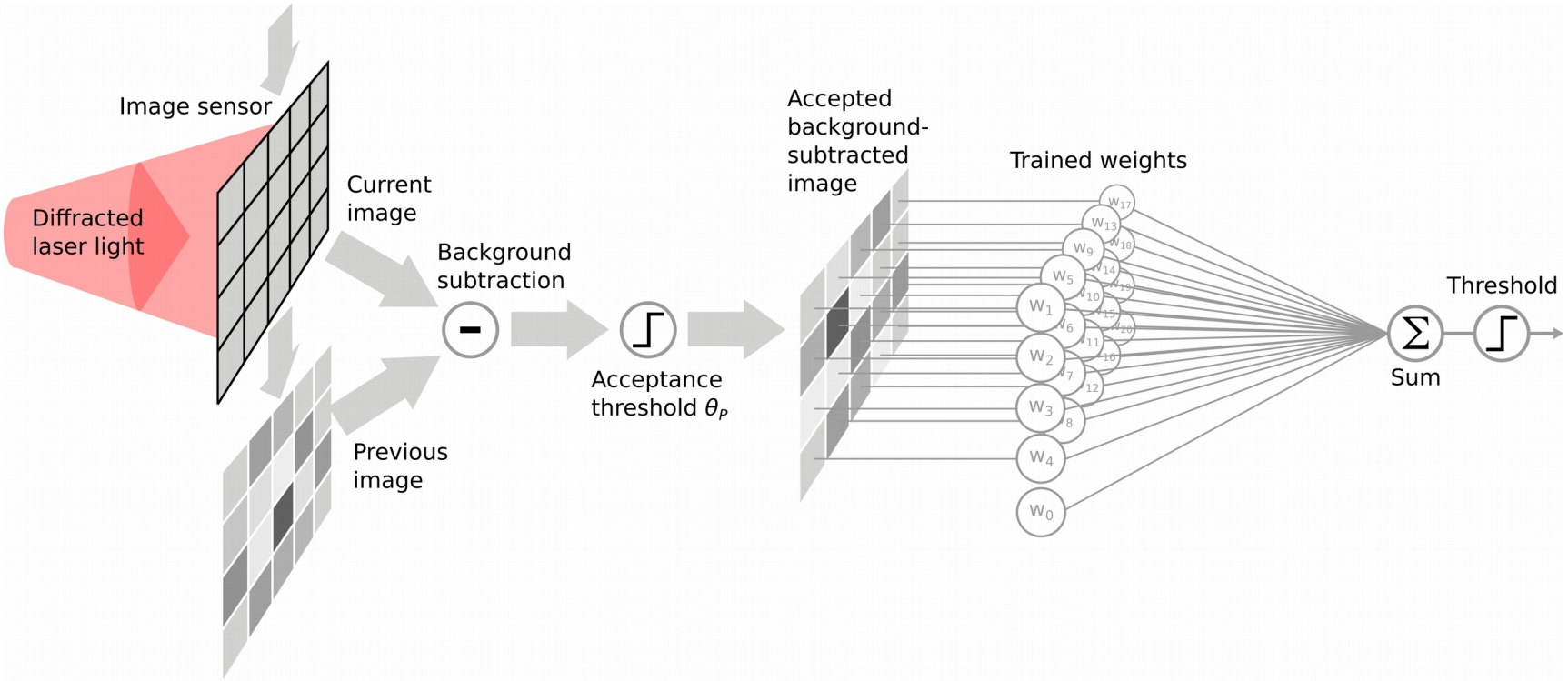
FINAL EXPERIMENT RESULTS

FINAL EXPERIMENT

- No focusing
- Large field of view
- Cheap and simple components



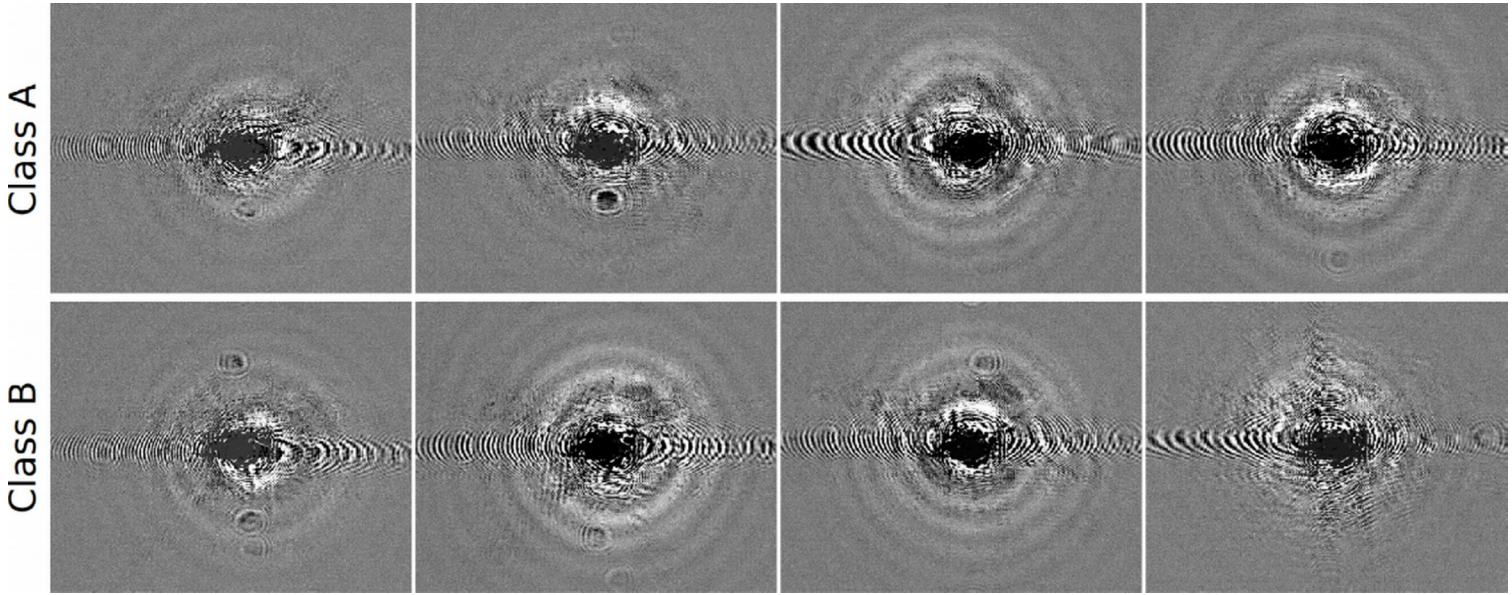
A SIMPLE, FAST AND VERSATILE CLASSIFIER



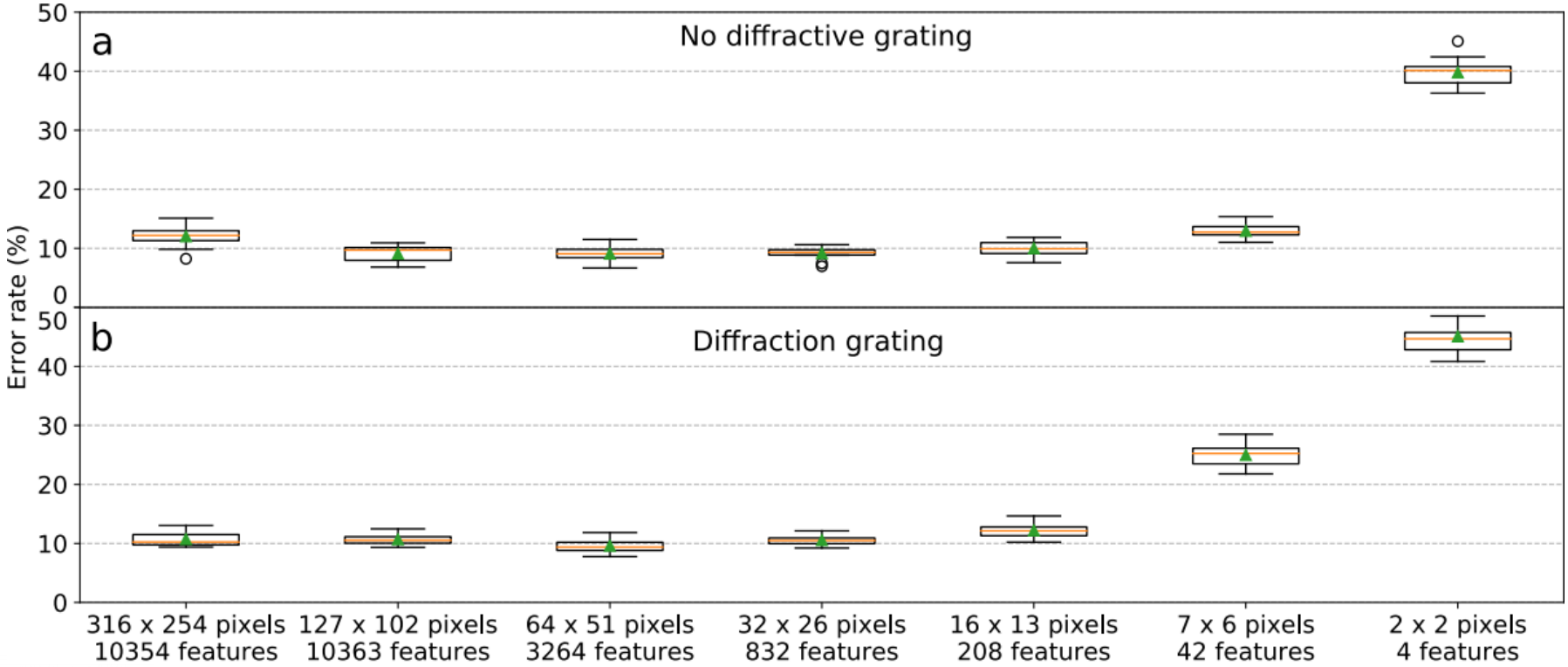
PATTERNS FROM DIFFERENT CLASSES

15µm

18µm



ERROR V.S. IMAGE RESOLUTION



GHENT UNIVERSITY imec

COMPARISON WITH OTHER WORKS

	Classification task	Classifier	Image resolution	Imaging method	Image FoV	Classification performance	Accelerator	execution time / particle	Meas. bias control
[1]	Beads with diameters of 7, 10 and 15 µm	CNN	21×21	Microscope	Centered and cropped	93.3% mAP	GPU	< 1 ms	Unreported
[2]	3 white blood cell (WBC) types	Rand. forest on extracted features	31×31	Lens-free - raw hologram	Unreported	96.8% accuracy	GPU	0.2 ms	Unreported
[3]	1 WBC type and an epithelial cancer cell	Deep CNN	Unreported	Time-stretch microscope	25 µm along channel	95.74% accuracy	GPU	3.6 ms	Unreported
	Beads with diameters of 15.2 and 18.6 µm (our work)	Linear (log. regression)	32×26	Lens-free - raw hologram	~ 300 µm along channel	> 90% accuracy	None	0.013 ms	Yes

Potentially close to ~100,000 cell/s

[1] Heo, Young Jin, *et al.* “Real-time image processing for microscopy-based label-free imaging flow cytometry in a microfluidic chip”. *Scientific Reports*, 2017.
[2] Cornelis, B., *et al.* “Fast and robust Fourier domain-based classification for on-chip lens-free flow cytometry.” *Optics Express*, 2018.
[3] Li, Yueqin, *et al.* “Deep cytometry: deep learning with real-time inference in cell sorting and flow cytometry.” *Scientific Reports*, 2019.

CONCLUSIONS

- A simple **linear classifier** can be applied to particle holograms to provide **ultra-fast classification** in label-free flow cytometry

- On **condition** that:
  - the **extreme learning machine** paradigm is considered
  - the **shortcut learning** due to varying measurement conditions is properly treated (we demonstrated a suitable methodology)

- The demonstrated approach is **simple** to employ, **versatile** and require few **cheap** components

CONCLUSIONS

- A simple **linear classifier** can be applied to particle holograms to provide **ultra-fast classification** in label-free flow cytometry
- On **condition** that:
  - the **extreme learning machine** paradigm is considered
  - the **shortcut learning** due to varying measurement conditions is properly treated (we demonstrated a suitable methodology)
- The demonstrated approach is **simple** to employ, **versatile** and require few **cheap** components

FUTURE PERSPECTIVES

- High-throughput with high-speed **event-based camera** (Muhammed Gouda in Neoteric project)
- Apply our method to **cell classification** (e.g. WBC)
- Apply method to **existent high-throughput imaging systems** (e.g. time-stretch microscopy) to enable online operations
- Can scattering layers improve classification in **single-pixel configuration**?

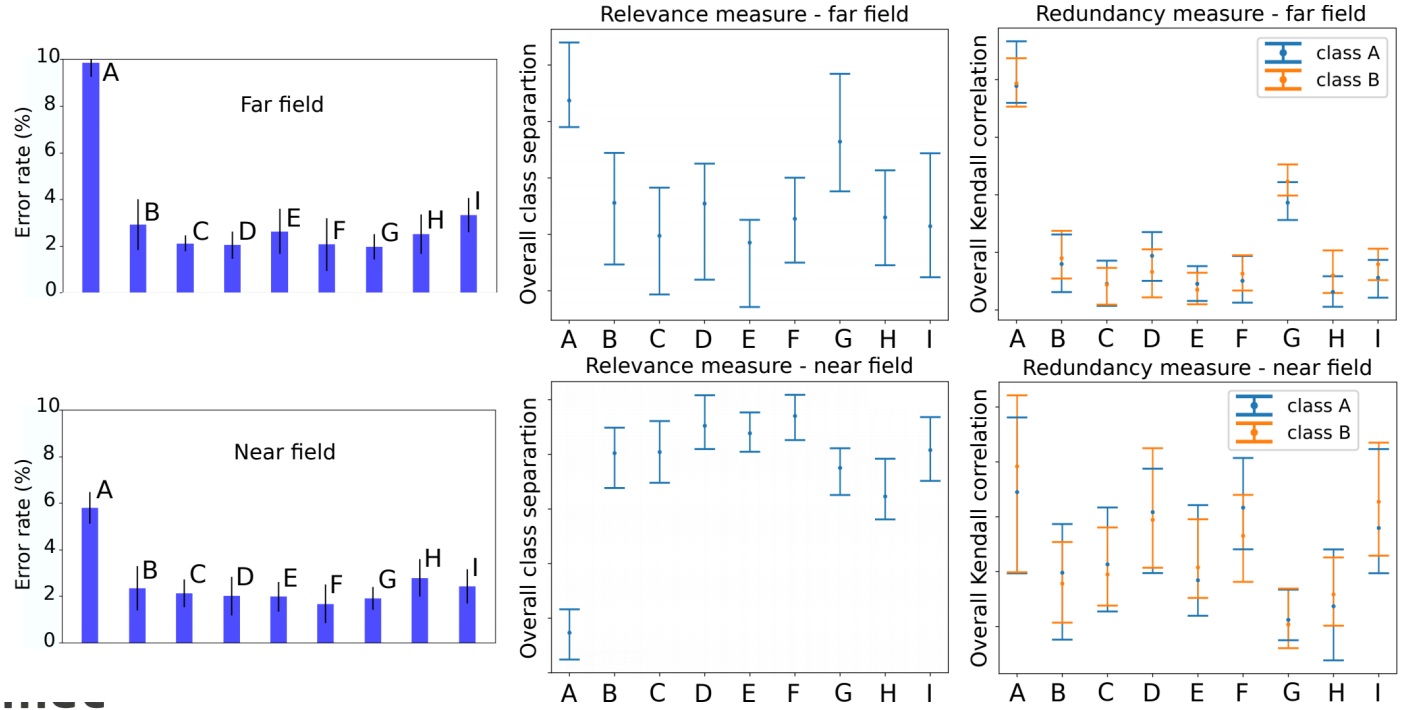
INTERPRETATION OF “DIMENSIONALITY EXPANSION”

Dimensionality expansion → Possible meanings:

1) more features per samples
2) more uncorrelated features
3) more relevant and uncorrelated features

redundancy
noise

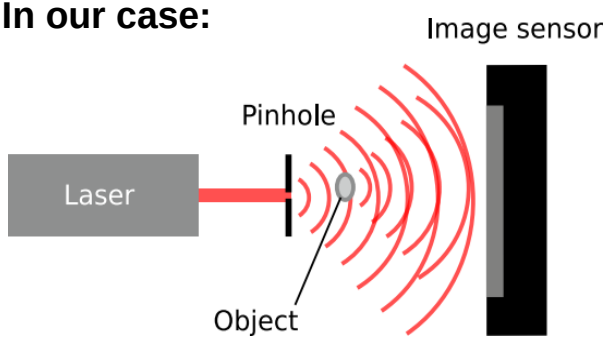
Enhanced linear separability



A SHORTCUT LEARNING EXAMPLE

Task to learn:
distinguish seabirds from crows in a picture

In our case:



$$h_i - h_{i-1} = \delta h_i^{\text{bkgr}} + \mathcal{N}(h_i^{\text{bkgr}}, C_i)$$

Nonlinear interaction between cell information and background
→ **need to act on measurements!**

Seabird examples

⋮

Crow examples

⋮

TRADE-OFFS

Complexity of the model

↑ dimensionality/parameters

↑ # required samples

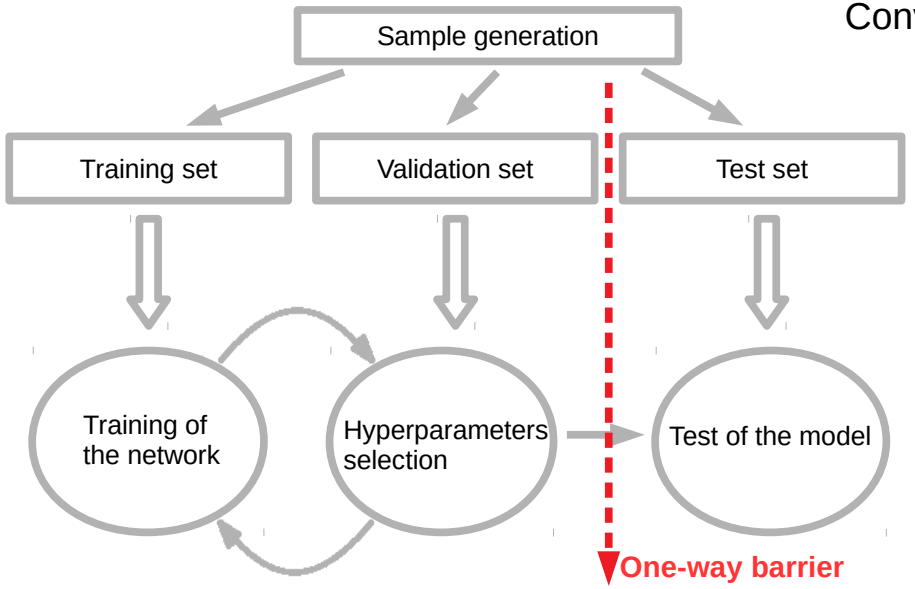
- Training hindered by local optimum points
- ↑ complexity ⇨ ↑ comp. cost

Number of samples ⇔ Computational cost

↓ # samples ⇨ ↑ # comp. cost

(e.g. k-fold cross-validation)

TREATMENT OF MEASUREMENT BIAS



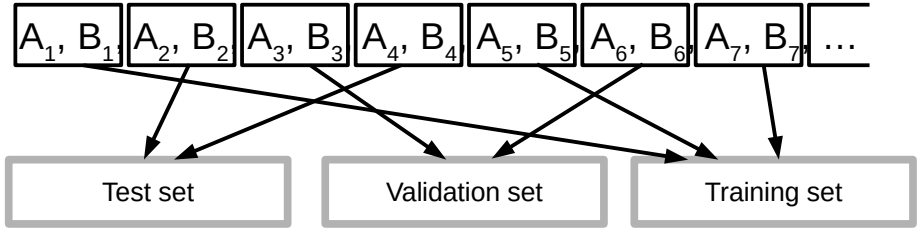
Conventional validation methods do not solve the problem

Measurement bias is an **elusive**, **two-fold** problem:

1) sidetracks the training algorithm → undermines learning
2) performance evaluation is also biased → test accuracy is inflated

Intertwined class measurements address both:

$A_1, B_1, A_2, B_2, A_3, B_3, A_4, B_4, A_5, B_5, A_6, B_6, A_7, B_7, \ldots$

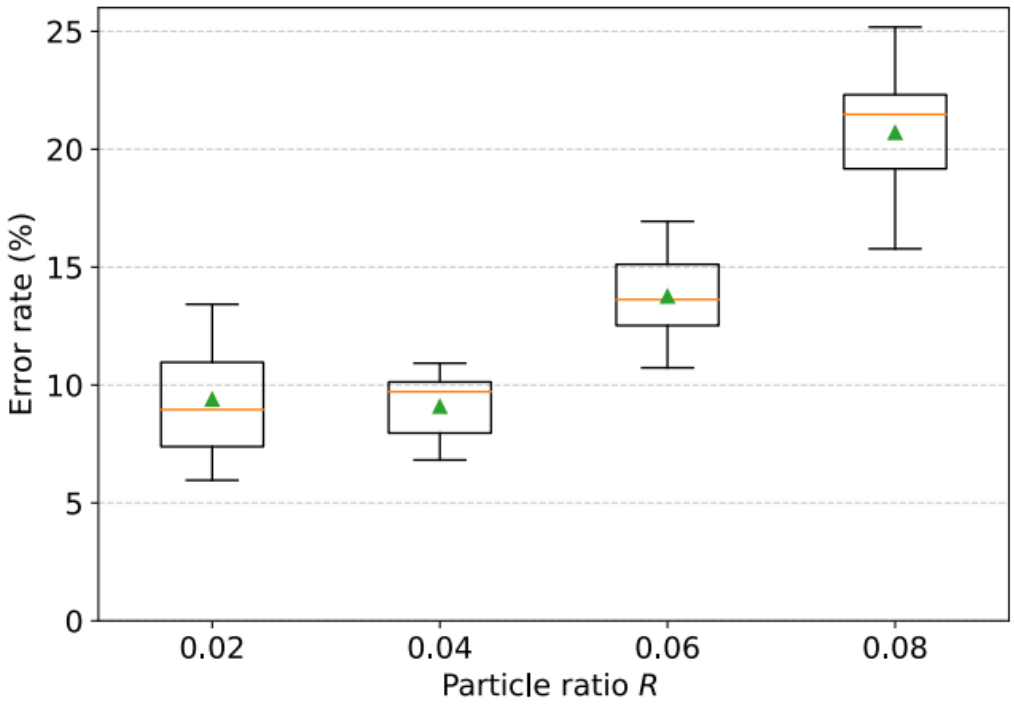


1. class - noise correlation is broken in training set
2. training, validation and test sets do not share the same measurement conditions

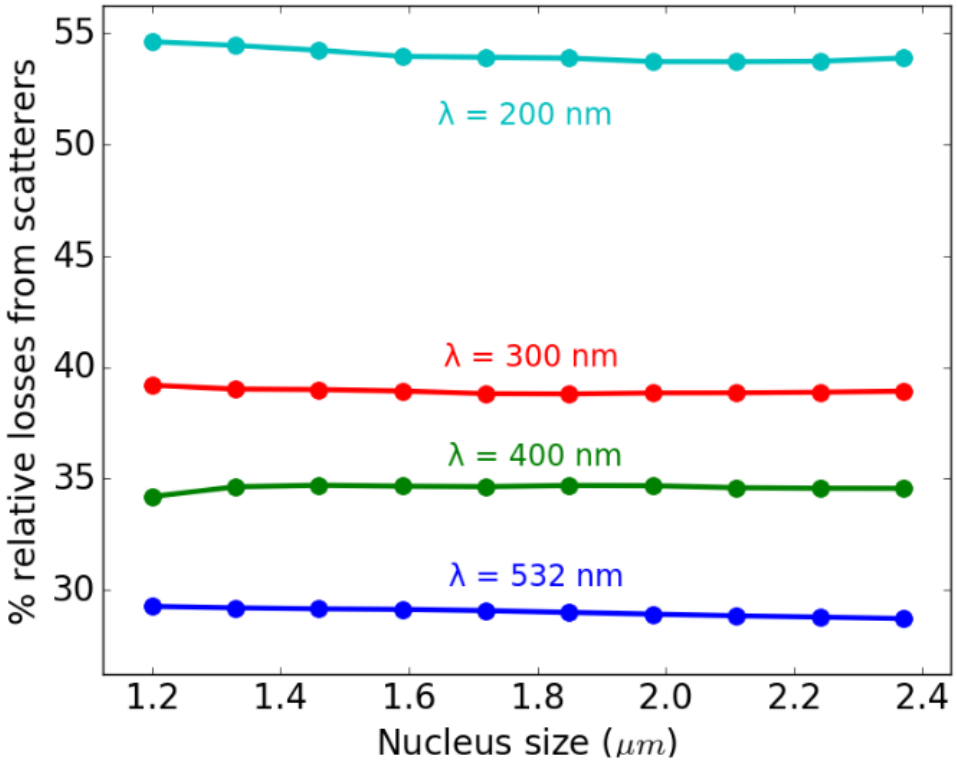
In Chapter 5:

- both aspects of meas. bias demonstrated
- removed by intertwined measurements

TRADE-OFF BETWEEN FIELD OF VIEW AND NUMBER OF SAMPLES

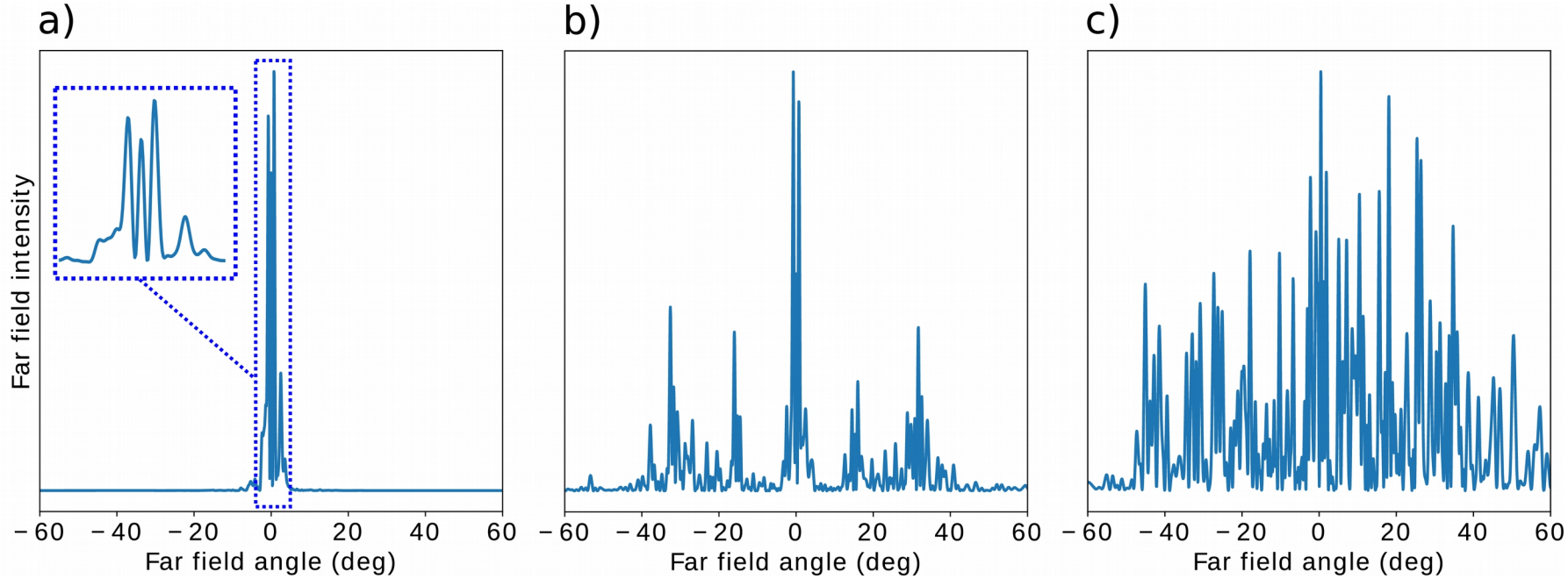


RELATIVE LOSSES WITH 4 SCATTERING LAYERS



Relative to the case without scatterers

INTENSITY IS MORE SPREAD OVER THE IMAGE SENSOR USING SCATTERERS

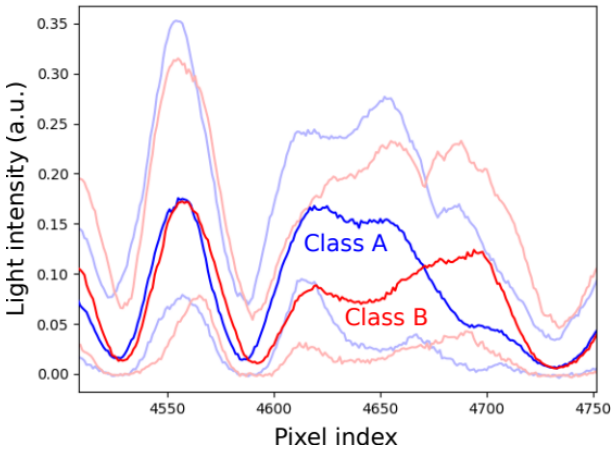


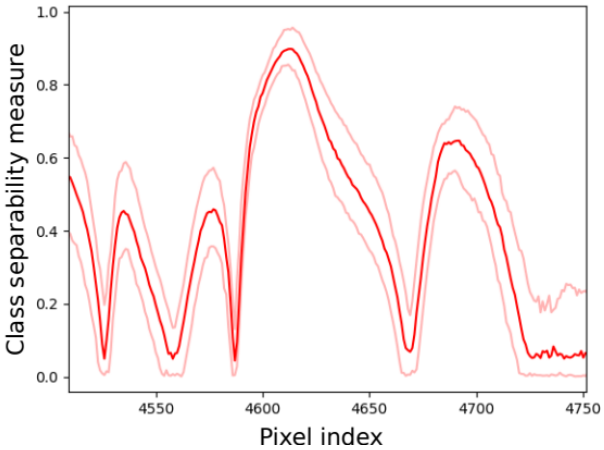
MANN-WHITNEY U STATISTIC AND KENDALL CORRELATION

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} S(X_i, Y_j),$$

with

$$S(X,Y) = \begin{cases} 1, & \text{if } Y < X, \\ \frac{1}{2}, & \text{if } Y = X, \\ 0, & \text{if } Y > X. \end{cases}$$





Correlation calculated on pixel pairs

$$\tau_A = \frac{n_c - n_d}{n_0}$$

$$n_0 = n(n-1)/2$$

$$n_c = \text{Number of concordant pairs}$$

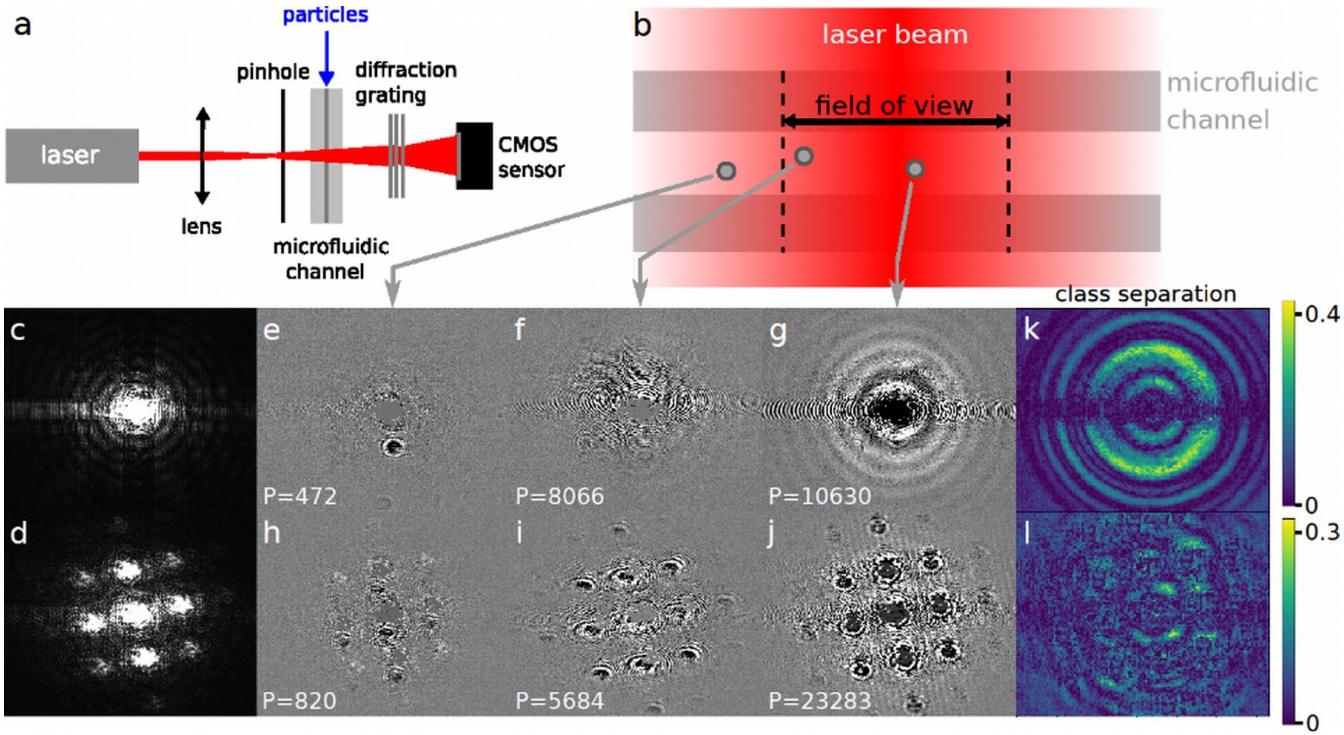
$$n_d = \text{Number of discordant pairs}$$

Source: Wikipedia

FIELD OF VIEW...

Possible improvements:

- higher background-to-noise ratio
- measure more samples
- explore scattering configuration on morphology-based classification task (e.g. WBC)
- partially automatized setup



INTERFERENCE PATTERN

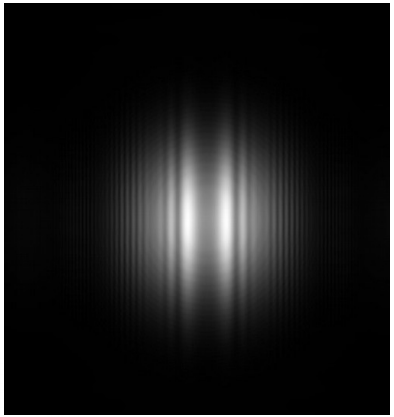
$$E(x,y,z) = \frac{e^{ikz}}{i\lambda z} e^{i\frac{\pi}{\lambda z}(x^2+y^2)} \mathcal{F}\left\{ E(x',y',0) e^{i\frac{\pi}{\lambda z}\left(x'^2+y'^2\right)} \right\}\Bigg|_{p=\frac{x}{\lambda z},\ q=\frac{y}{\lambda z}}$$

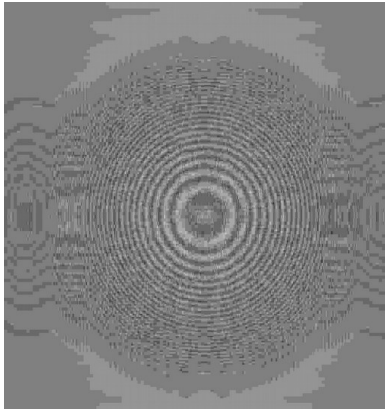
Fresnel diffraction (near field)

$$U(x,y,z) \propto \hat{f}\left[A(x',y')\right]_{f_x f_y}$$

Fraunhofer diffraction (far field)

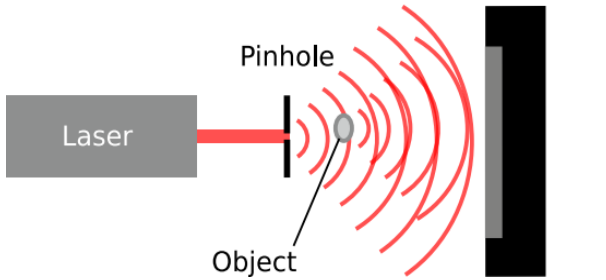
Mie scattering is most suitable when the micorparticle dimension is comparable with the wavelength



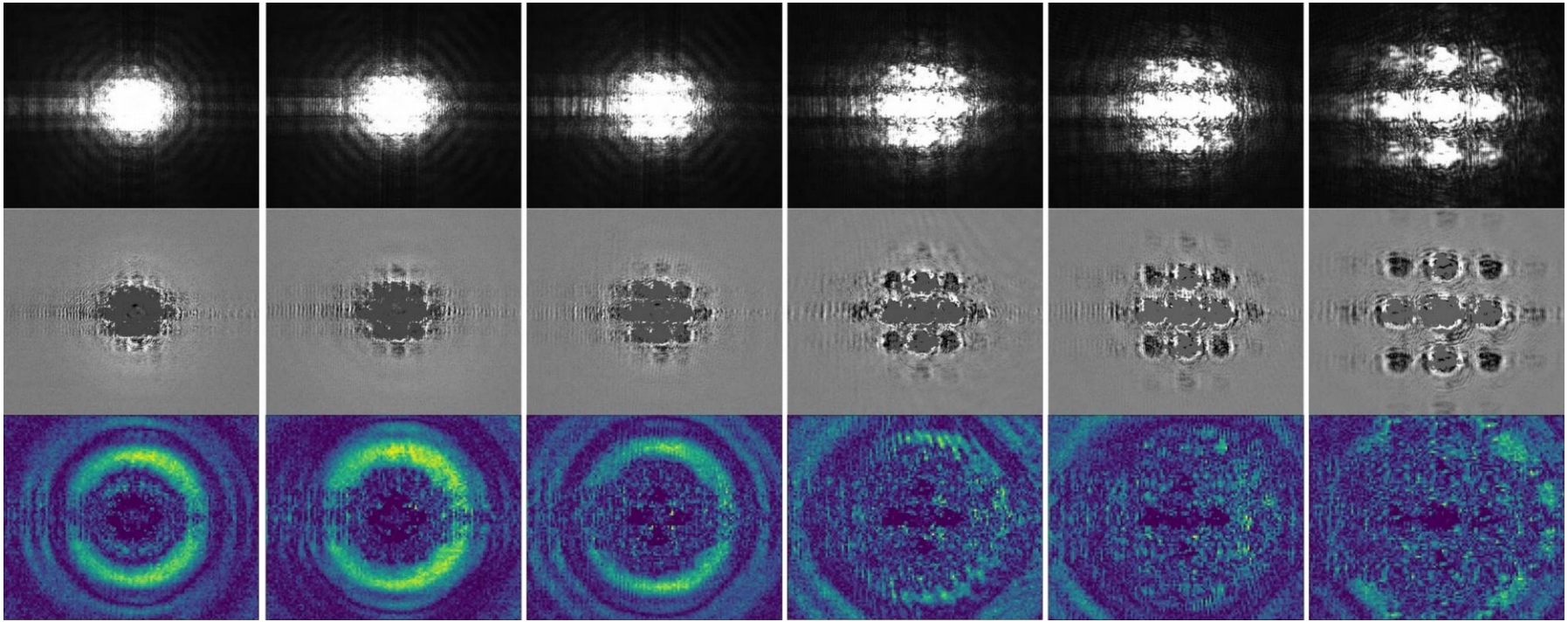


Emmanuel Gooskens
UGent

Inline digital holographic microscopy



DEPENDENCY ON CAMERA POSITION



FOV ESTIMATION

from k particles in the FoV can be considered as the Poisson process describing the occurrence of k events t_{in}, with a time rate R_f, in a time interval $\tau + \text{FoV}/v$, with probability:

$$Pr(k, \tau + \text{FoV}/v, R_f) = \frac{[R_f(\tau + \text{FoV}/v)]^k}{k!} e^{-R_f(\tau + \text{FoV}/v)} \tag{5.4}$$

In our case $\tau = 29\,\mu\text{s}$ and we can calculate R_f by multiplying the flux rate (0.2 ml/min) by the estimated particle concentration, which depends on the particle class (1.6×10^4 and $0.91 \times 10^4 \frac{\text{particles}}{\text{ml}}$ respectively for class A and B) since the mixtures have a common solid content volume. Note that we are assuming that the number of particles that remain stuck somewhere before reaching the illumination area is negligible w.r.t. the total number of passing particles. Therefore, even if we deem this assumption sufficiently true in our case, we should keep in mind that the estimated R_f is more an upper limit for the true particle flow rate. From the next calculation steps it will be evident that this implies that we will obtain a lower limit estimate of the true FoV. To provide an example calculation, assuming a reasonable FoV= 100 µm, respectively for classes A and B we obtain (keeping 2 significant digits): $Pr_A(k = 0)$ = 0.98, $Pr_B(k = 0) = 0.99$, $Pr_A(k = 1)$ = 0.017, $Pr_B(k = 1) = 0.0098$, $Pr_A(k = 2)$ = 0.00016, $Pr_B(k = 2) = 0.000048$. These through the microfluidic channel (statistical independence). The particle ratio R can be estimated by $R = 1 - Pr(0, \tau + \text{FoV}/v, R_f)$, with reference to equation (5.4). Thus, by inverting it, we can finally estimate the FoV corresponding to a chosen value of R:

$$\text{FoV} = -\frac{\ln(1 - R)v}{R_f} - \tau v \tag{5.5}$$

For each chosen value of R and for each particle class, we report in Table 5.3 the number of classification samples (accepted images) and the FoV estimates. The corresponding estimated FoV is quite large: ≈ 0.3 mm. It should also be stressed that, as a consequence of our choice of having a single threshold θ_P for both classes and for training and testing, the FoV was class-dependent.

No diffractive layer

Particle rate	# accepted images		Field of view (mm)	
	class A	class B	class A	class B
0.02	1427	2108	0.09	0.25
0.04	4008	3067	0.27	0.37
0.06	6452	4120	0.45	0.51
0.08	7954	6051	0.56	0.76

Frame rate ~ 138 fps, exposure time = 20 us, ~5.5 p/s

EXAMPLE ARTICLES

Lippeveld, Maxim, Carly Knill, Emma Ladlow, Andrew Fuller, Louise J. Michaelis, Yvan Saeys, Andrew Filby, and Daniel Peralta. "Classification of human white blood cells using machine learning for stain-free imaging flow cytometry." Cytometry Part A 97, no. 3 (2020)

- Proper ground truth with manual gating
- Deep learning does not outperform feature engineering
- 8 different types of WBC, but also 3 types classification
- Accuracy < 90%

Tang, Rui, Zunming Zhang, Xinyu Chen, Lauren Waller, Alex Ce Zhang, Jiajie Chen, Yuanyuan Han, Cheolhong An, Sung Hwan Cho, and Yu-Hwa Lo. "3D side-scattering imaging flow cytometer and convolutional neural network for label-free cell analysis." APL Photonics 5, no. 12 (2020)

- Label-free using light sheet and side scattering.
- 92% accuracy WBC classification
- Ground truth with manual gating