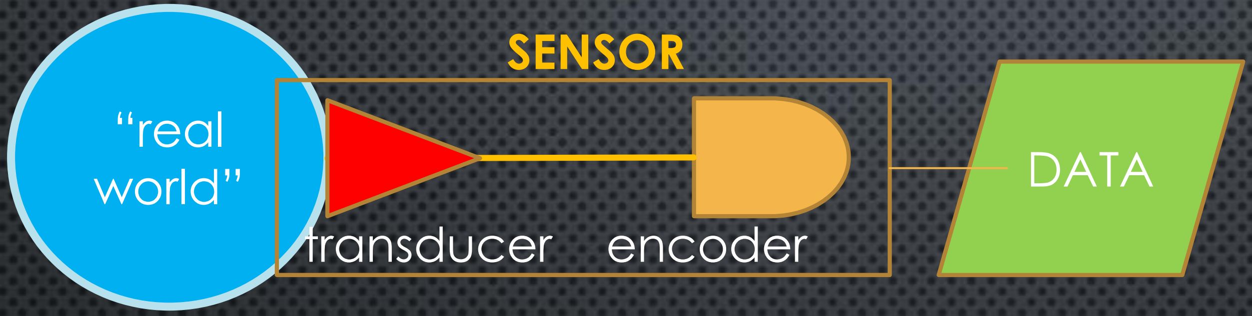
# BIG DATA IN REAL TIME: FAST AND FURIOUS

ABDALLAH[1], G. AIELLI[1], R. CARDARELLI[1], M. MANCA[3]

[1] UNIVERSITY AND INFN OF ROMA TOR VERGATA, ITALY; [2] CERN

# DATA OR BIG DATA?

- Sensors produces DATA →
  - A sensor is whatever is able to produce an observable encoding events in a given format




- Too BIG for what?
  - Sensors produce data faster than we can store
  - We need to interact with data to take decisions in real time
- BIG is a relative concept – not just a matter of volume
  - More than the absolute volume the saliency matters. Model dependent
  - Successful data reduction model ? → Thermodynamics! A thermometer is a smart sensor

# BIG DATA: NOT ALL THE SAME:

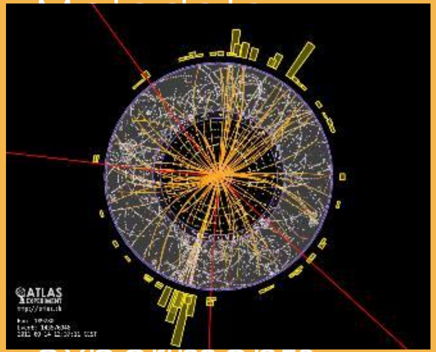
Data model complexity

## Structured

Da… …ses:
Homo… …ous and
fixed … …t data
with a … …d range
v… …ity:
- A ph… …directory
- me… …ements

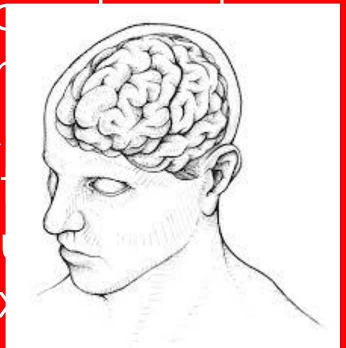

## Multi-Structured /Hybrid

- 
- 
- 
- … experiments




## Free-structured

- So…
- En…
- A…
- Li…
- Hu… ex…
- Scientific models



- Data model fixed
- Small parametric space to represent the data
- Prevalent deduction based analysis

- Data model evolving with data
- Variable parametric space
- Data driven induction and deduction

- Data model evolving with experience
- morphing parametric space
- Analogy based induction and deduction

# DATA INTENSITY AND DATA RATE: FAST AND FURIOUS

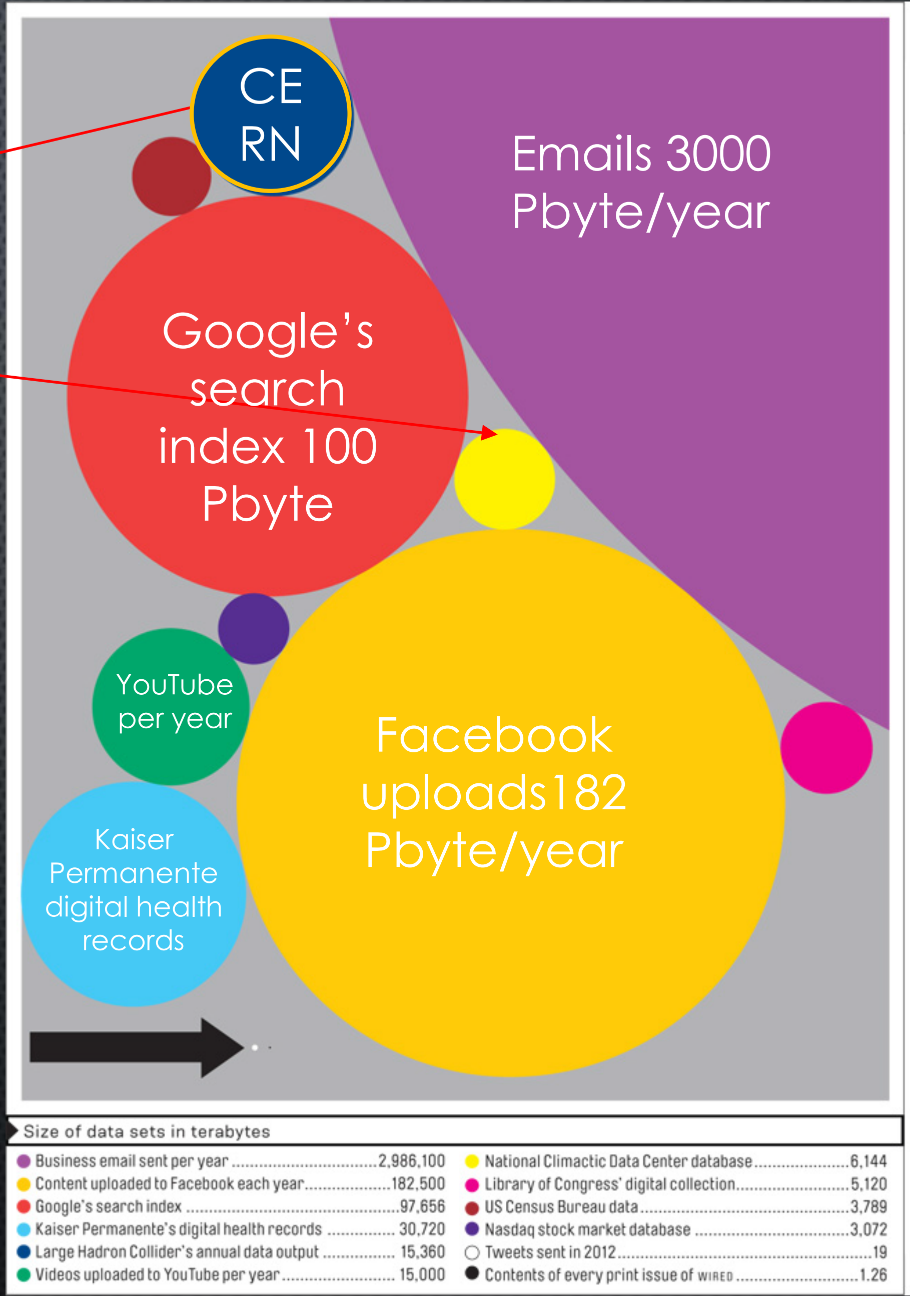
[HEP may have something to say!]

- THE DATA VOLUME CAN BE DRIVEN BY SCATTERED OR CONCENTRATED SENSOR SYSTEMS
  - GEOGRAPHICALLY SCATTERED DATA SOURCES → LOW DENSITY MANAGED THROUGH DATA LOGISTICS NETWORKS. THE REMOTE TERMINALS ARE SMART FRONT END
- CONCENTRATED SENSORS EACH PRODUCING HIGHER THAN MANAGEABLE DATA FLOW? A PARTICLE PHYSICS EXPERIMENT AT LHC FOR EXAMPLE…
  - SEARCH FOR A NEEDLE IN THE STRAW
  - NEED FOR SELECTION AT SENSOR LEVEL → TRIGGER CONCEPT
  - FAST AND SMART PROCESSING TO TAKE A FAST DECISION IN REAL TIME, AT NS SCALE…, FAST AND FURIOUS

For the HEP people this is a MUST since ever.
**Why all this hurry??**
Well, to avoid an experiment running 100 years….

# REALLY BIG DATA

- CERN "CLASSIC" BIG DATA 15 PBYTE/YEAR
- THIS IS AFTER THE TRIGGER SELECTING $1/(4*10^4)$
- THE REAL DATA ARE THEN 600 EXA BYTE/YEAR
- 19 PBYTE FROM A SINGLE SOURCE



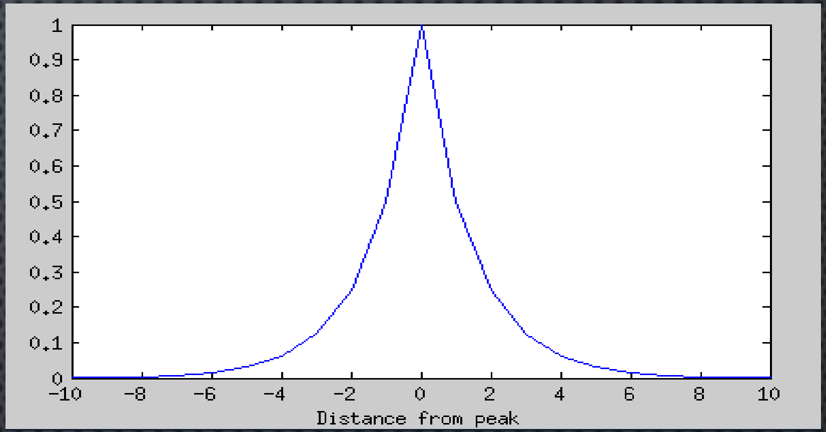

## MUST BE TREATED AT THE SOURCE

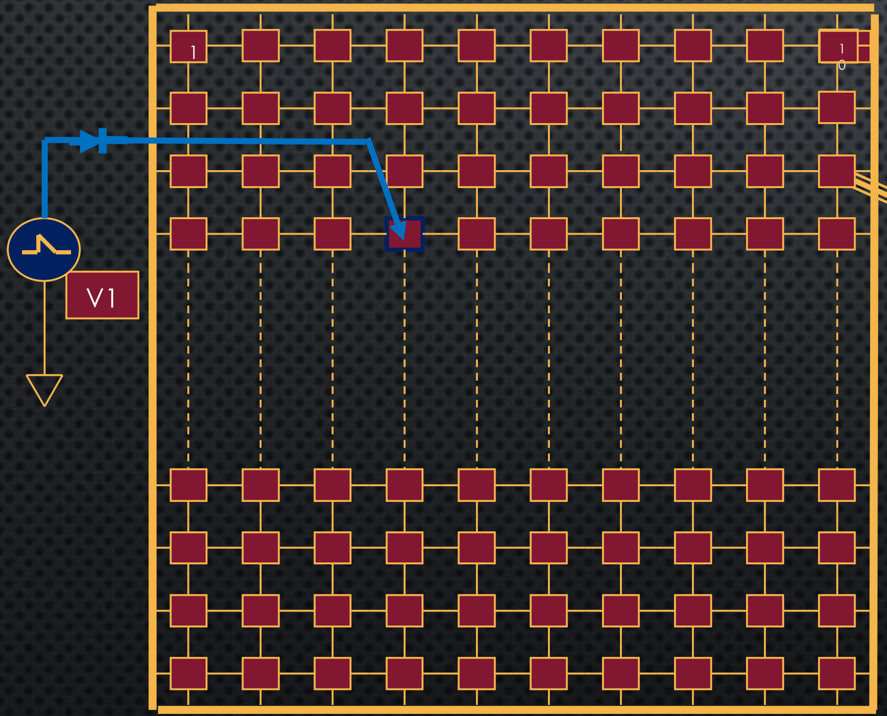
# HEP COMMUNITY AWARENESS AT THE DAWN OF ICT ERA

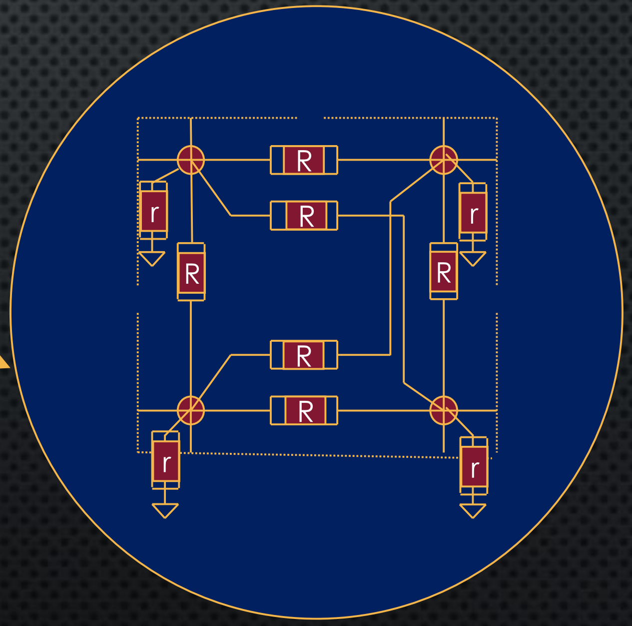
- In the '80 a big effort started to solve the problem of fast pattern recognition and neural networks
  - Ristori et al. Associative Memories (AM chip)
  - Darbo et al. Contiguity processor
  - AA.VV. Start of studies on neural network algorithms for tracking
  - AA.VV. Fuzzy logic
- Modern tracking triggers would like to:
  - Suppress the noise
  - Reconstruct the track elements
  - Classify meaningful events through topology
- A regression is the best way to evaluate data against a model→ computation complexity

# WEIGHTING RESISTIVE MATRIX: LET THE PHYSICS COMPUTE

- AN INTERCONNECTED REPETITIVE RESISTIVE NETWORK SUBJECTED TO AN ELECTROSTATIC POTENTIAL PROVIDES THE NECESSARY TOOLS
  - THE DIFFUSION POTENTIAL AS A NOTION OF DISTANCE
  - THE SUPERPOSITION PRINCIPLE ENSURES THE CORRELATION

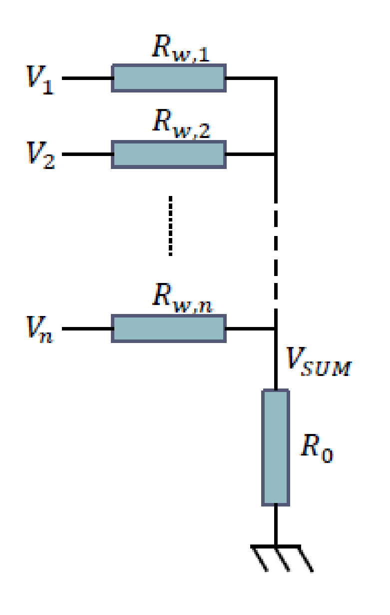


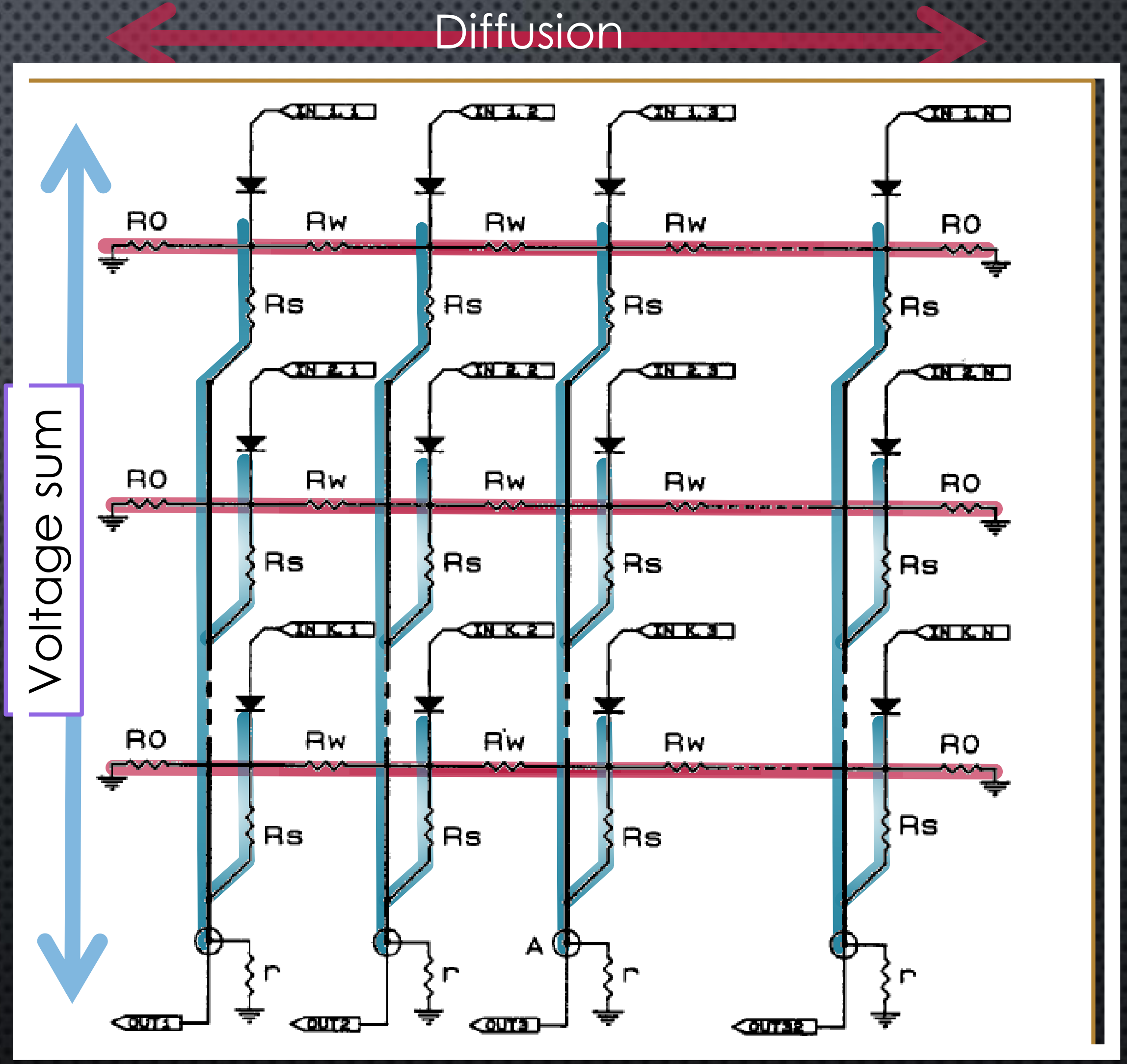
# THE WRM SCHEMA – A SIMPLE 1D IMPLEMENTATION

- From: R. Cardarelli et al.; "On a very fast topological trigger"; NIM A324(1993) 253-259
- A virtually infinite resistive network, where the ratio Rw/Rs determines the exponential voltage diffusion rate



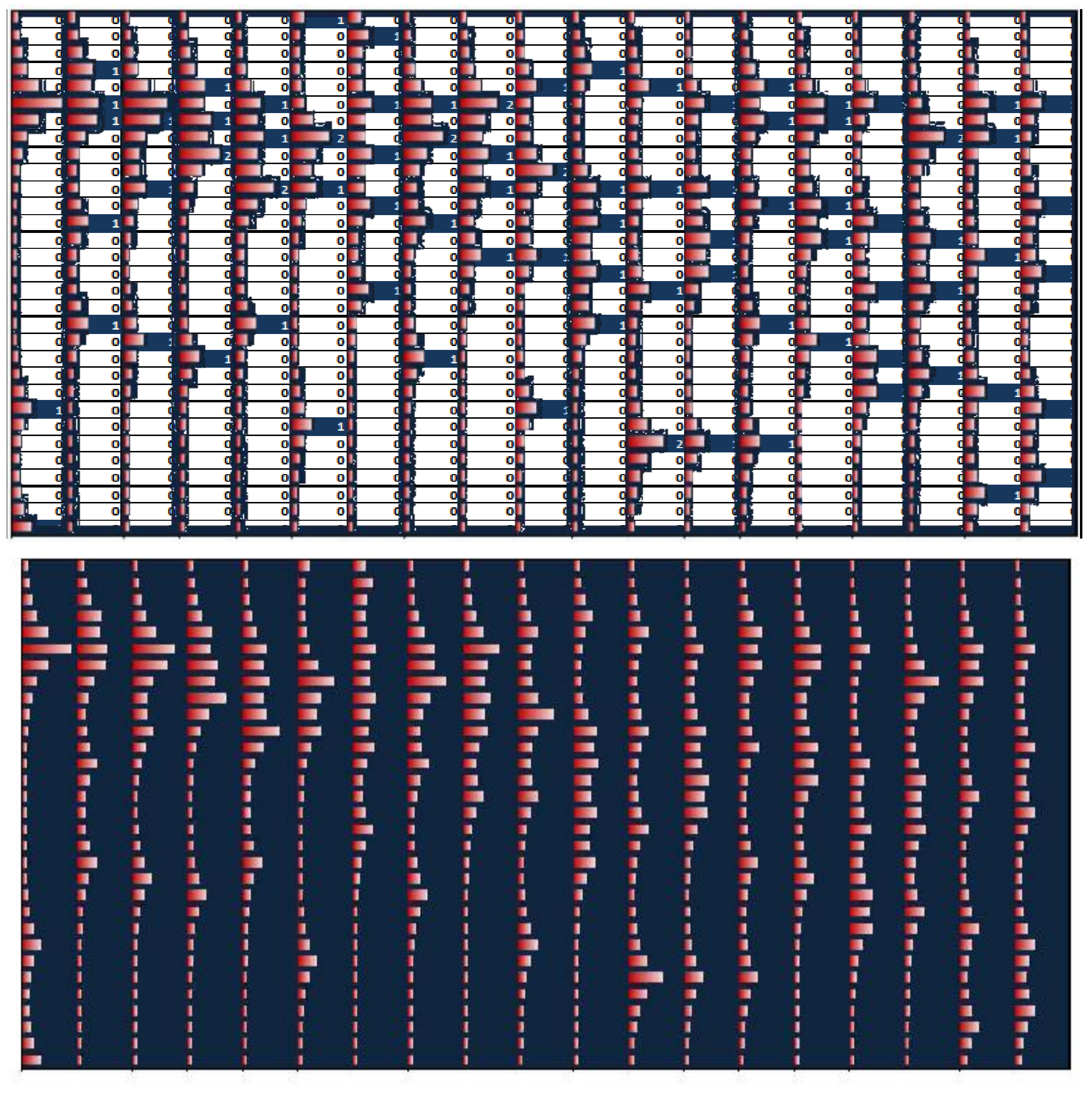

*If* $R_0 \ll \min(R_{w,1}, \dots, R_{w,n})$

$$\Longrightarrow V_{SUM} = w_1 V_1 + w_2 V_2 + \dots + w_n V_n$$



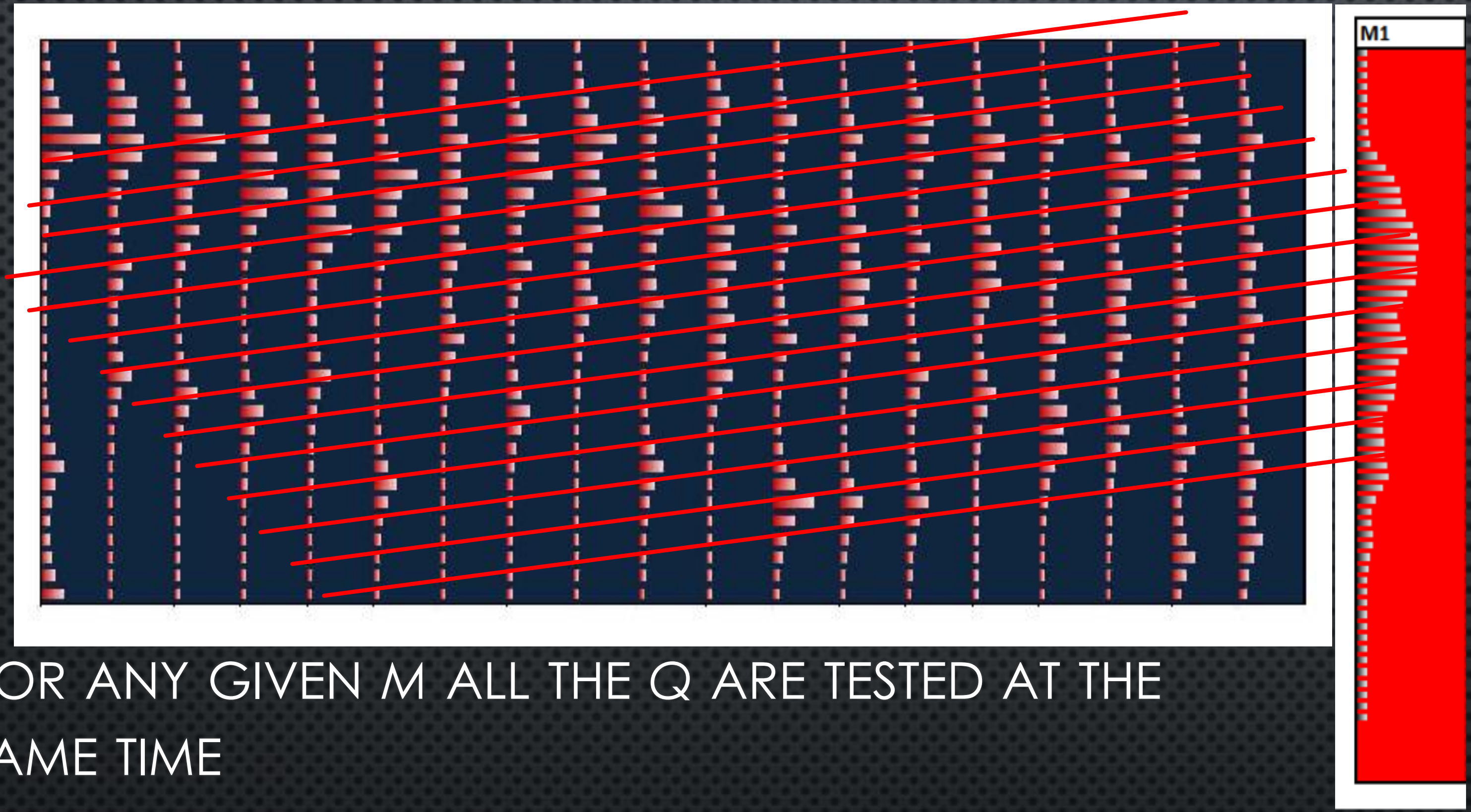
# WRM EXERCISE: 3 TRACK VERTEX WITH NOISE

- UPPER IMAGE IS THE DIGITAL INPUT TO THE WRM.
- CAN BE ANY DIGITAL READOUT DETECTOR



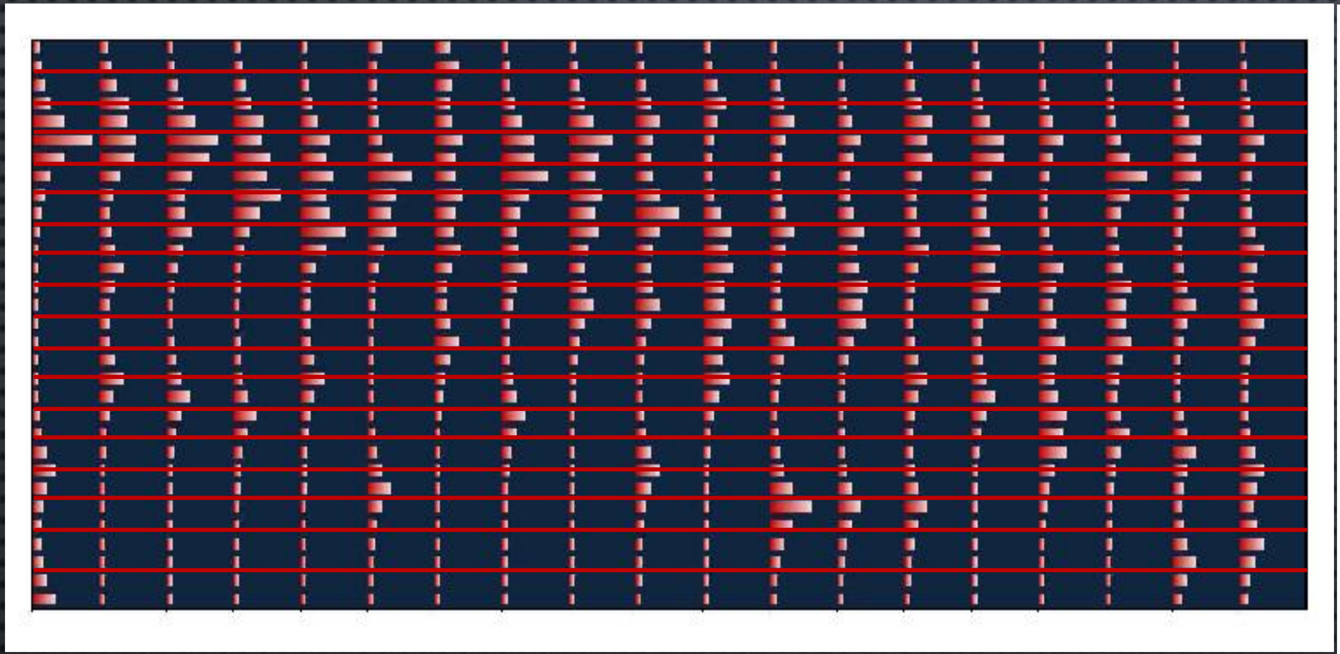
- LOWER IMAGE IS THE CORRESPONDING POTENTIAL DISTRIBUTION PRODUCED BY THE WRM.

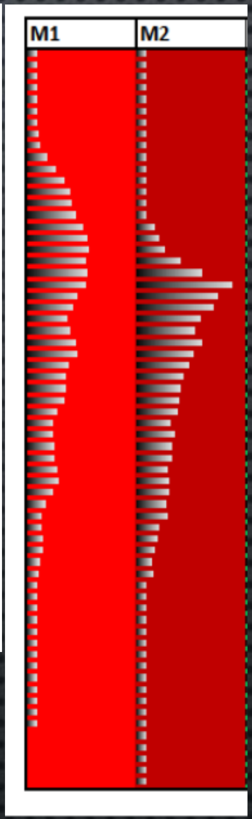
# STRAIGHT LINE FIT: Y=MX+Q




- FOR ANY GIVEN M ALL THE Q ARE TESTED AT THE SAME TIME

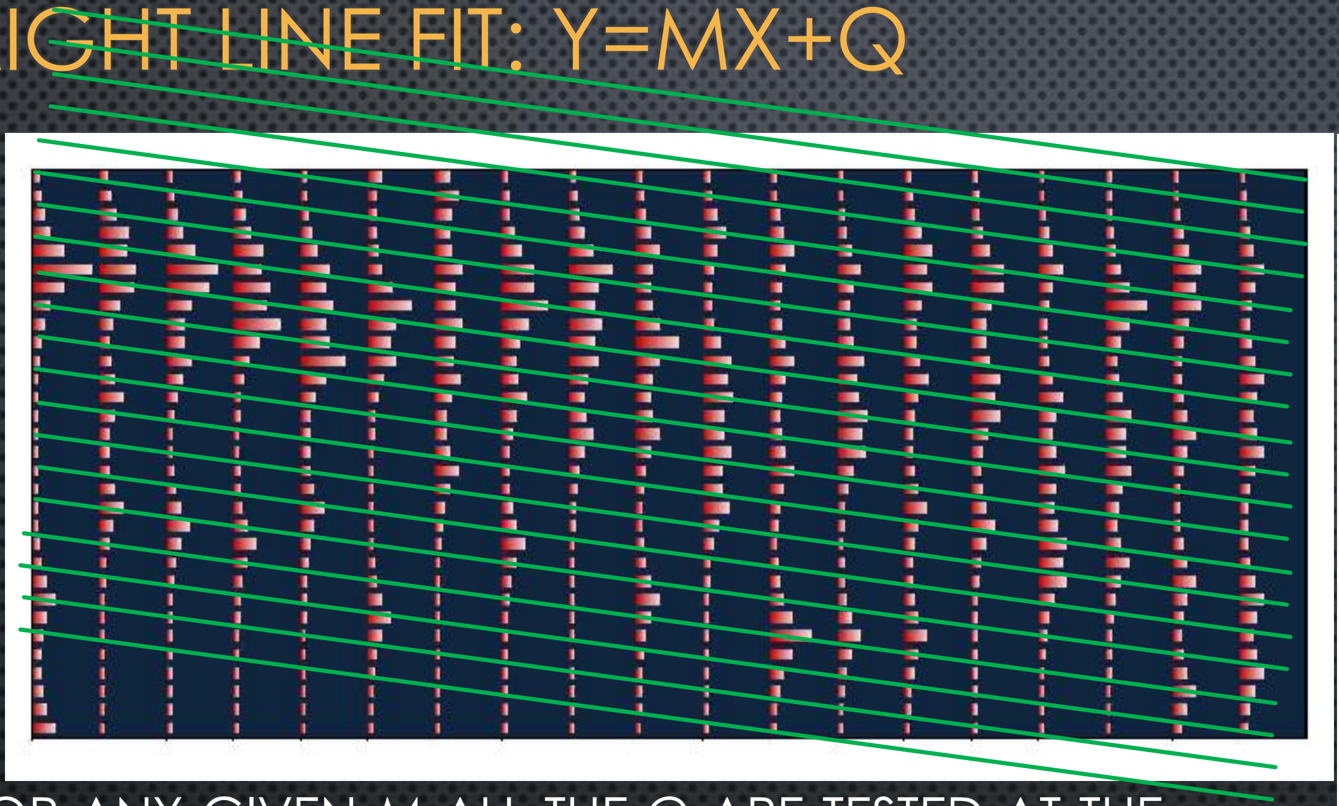
# STRAIGHT LINE FIT: Y=MX+Q

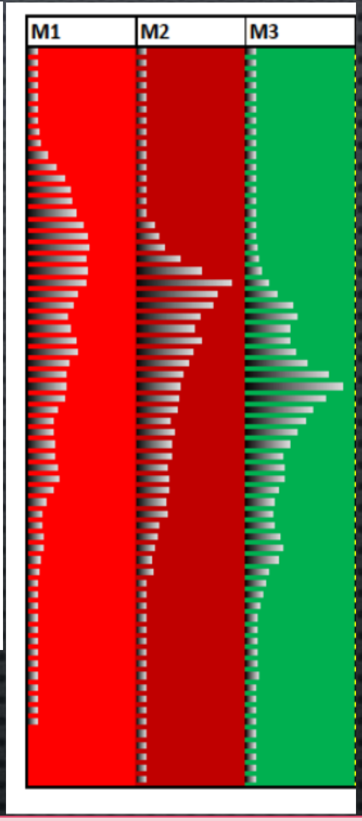





- For any given m all the q are tested at the same time
- A good m,q choiche produce a sharper and higher peak in the parameter space

# STRAIGHT LINE FIT: Y=MX+Q

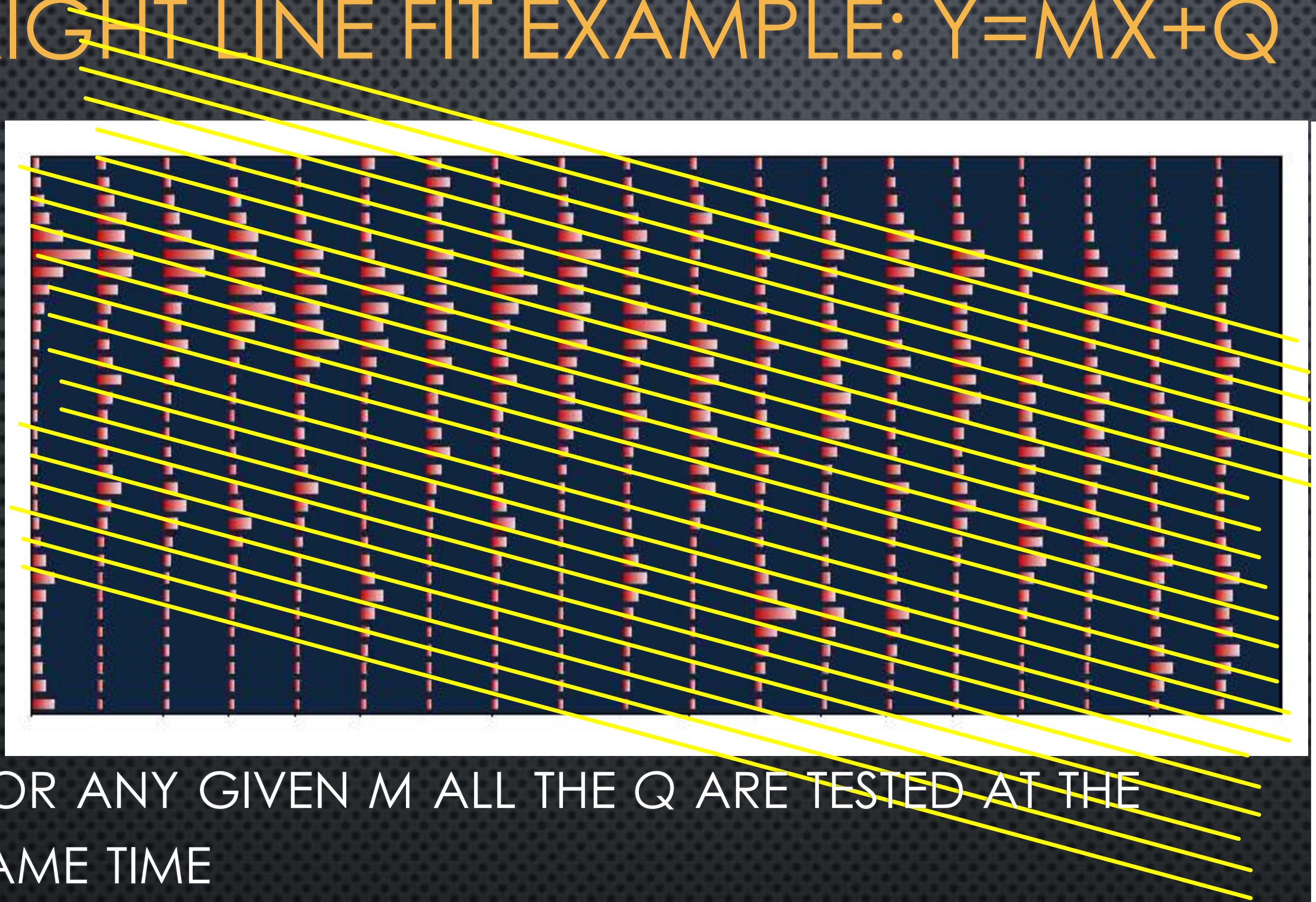





- FOR ANY GIVEN M ALL THE Q ARE TESTED AT THE SAME TIME
- A GOOD M,Q CHOICHE PRODUCE A SHARPER AND HIGHER PEAK IN THE PARAMETER SPACE

The test can be done in parallel or the data are feed to the matrix by a circular buffer

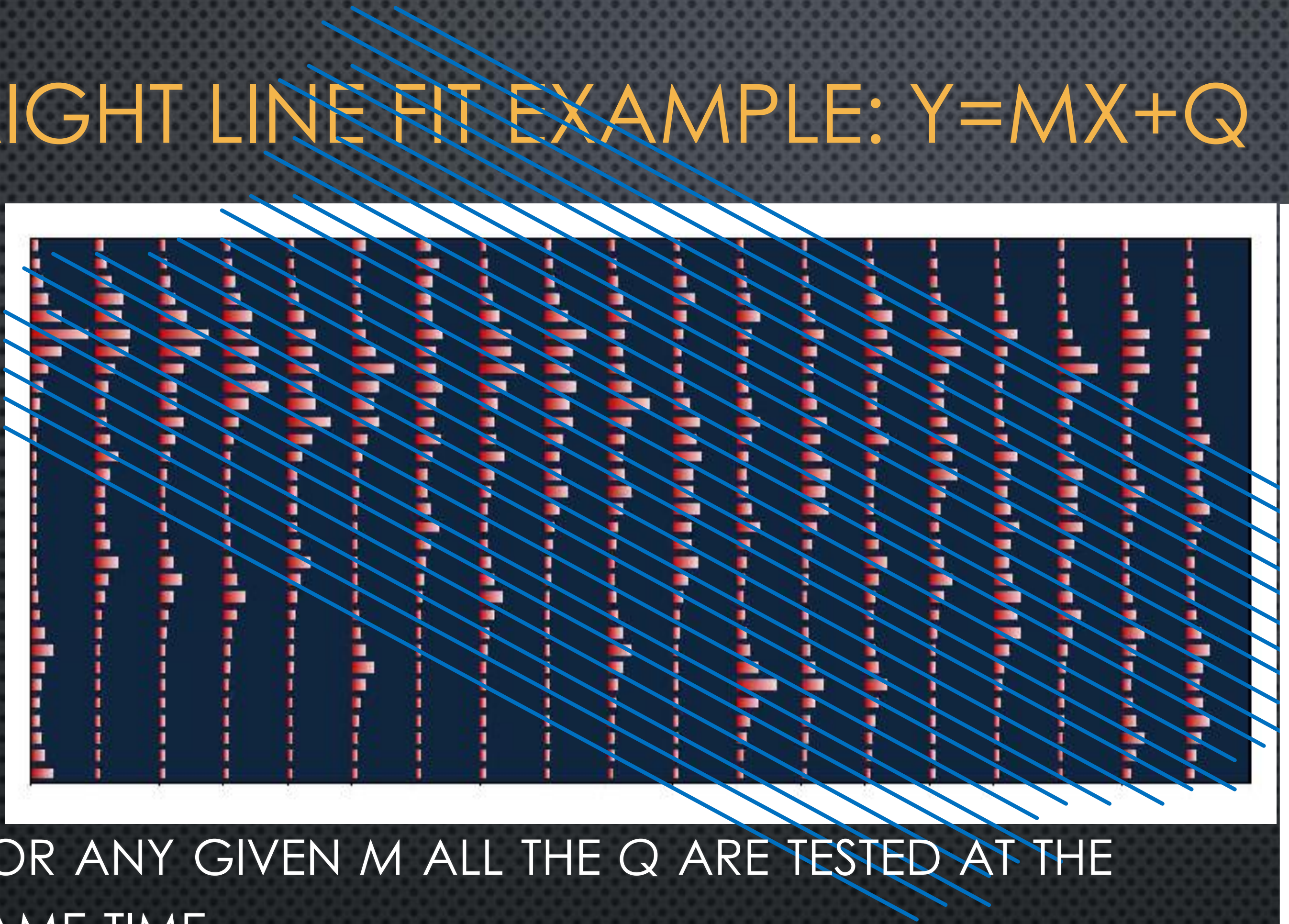
# STRAIGHT LINE FIT EXAMPLE: Y=MX+Q

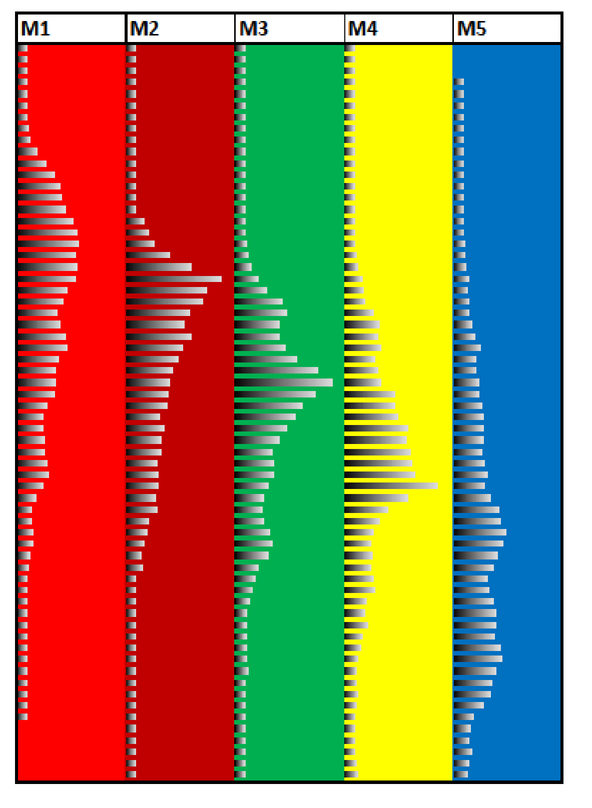



- For any given m all the q are tested at the same time
- A good m,q choiche produce a sharper and higher peak in the parameter space

The test can be done in parallel or the data are feed to the matrix by a circular buffer

# STRAIGHT LINE FIT EXAMPLE: Y=MX+Q

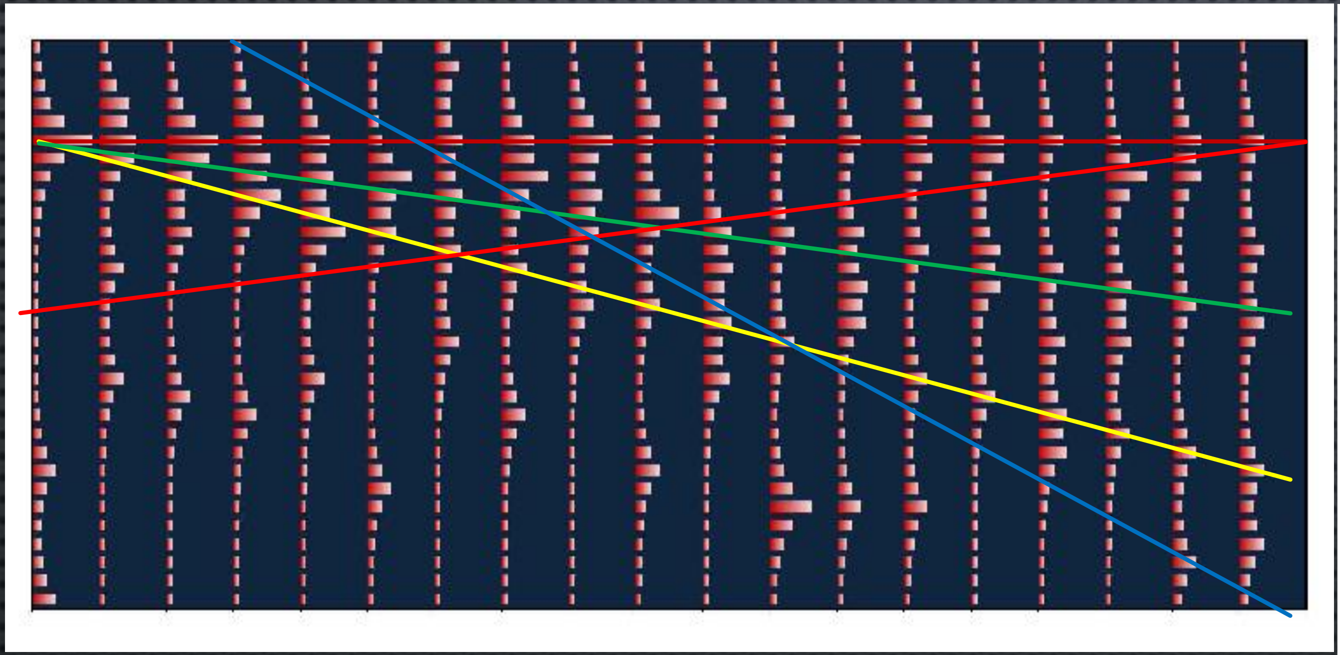


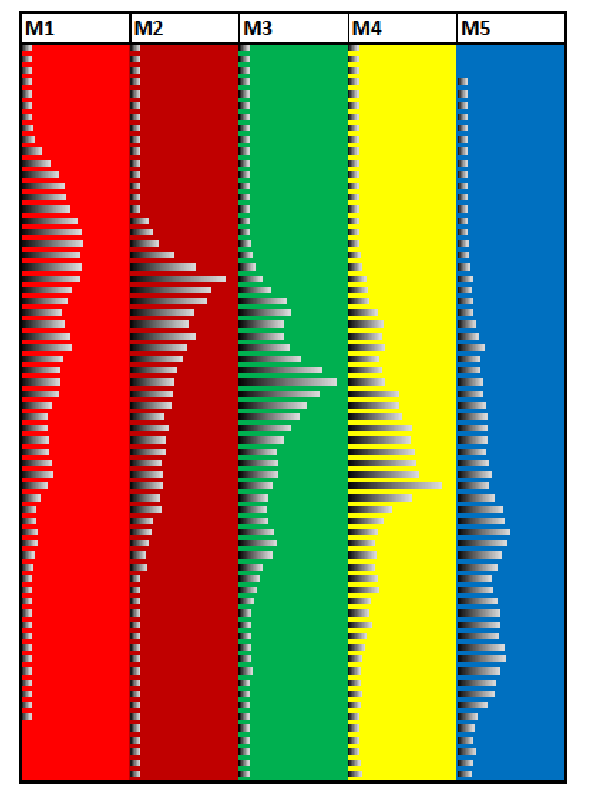



- FOR ANY GIVEN M ALL THE Q ARE TESTED AT THE SAME TIME
- A GOOD M,Q CHOICHE PRODUCE A SHARPER AND HIGHER PEAK IN THE PARAMETER SPACE
- IN THE OTHER CASES THE DISTRIBUTION IS FLAT

The test can be done in parallel or the data are feed to the matrix by a circular buffer

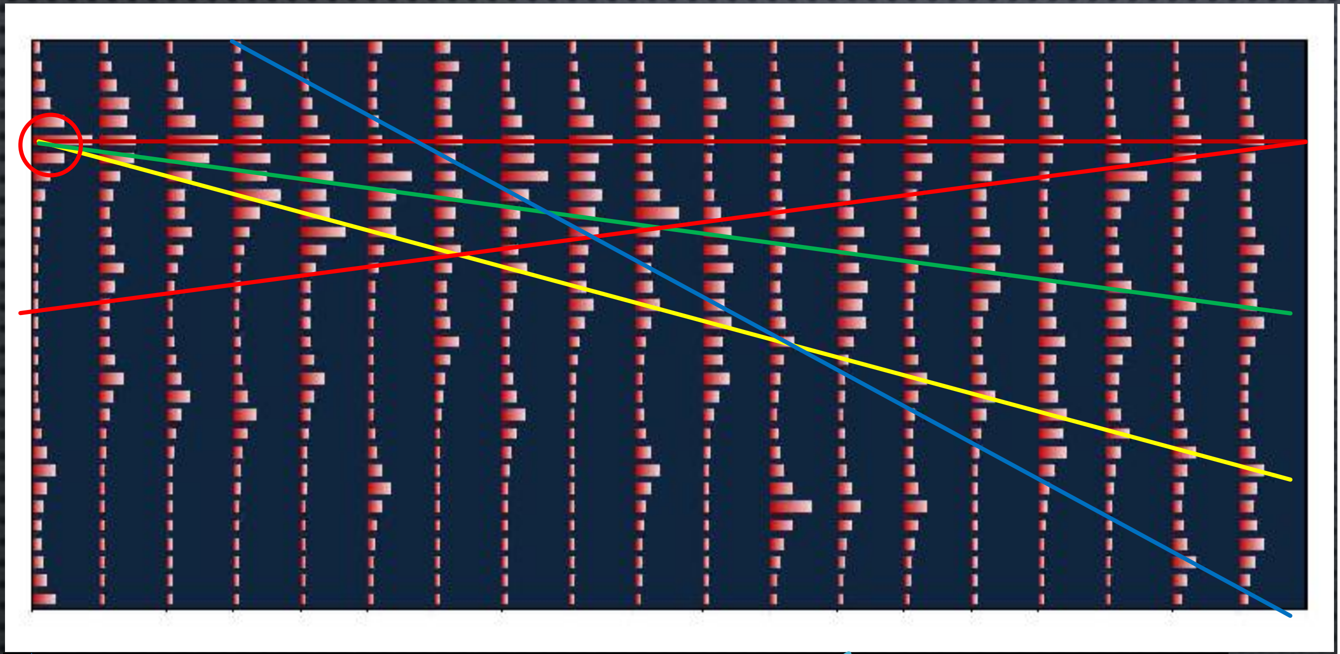
# VERTEX IDENTIFICATION EXAMPLE

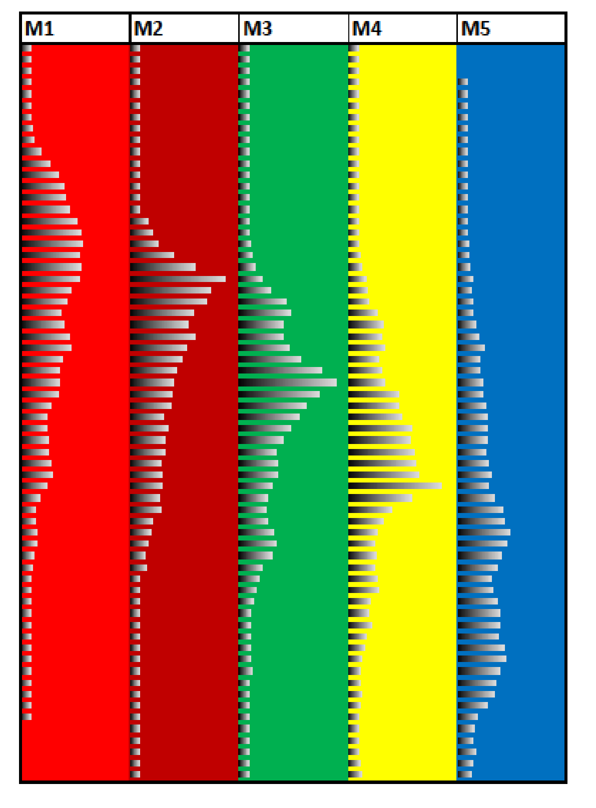


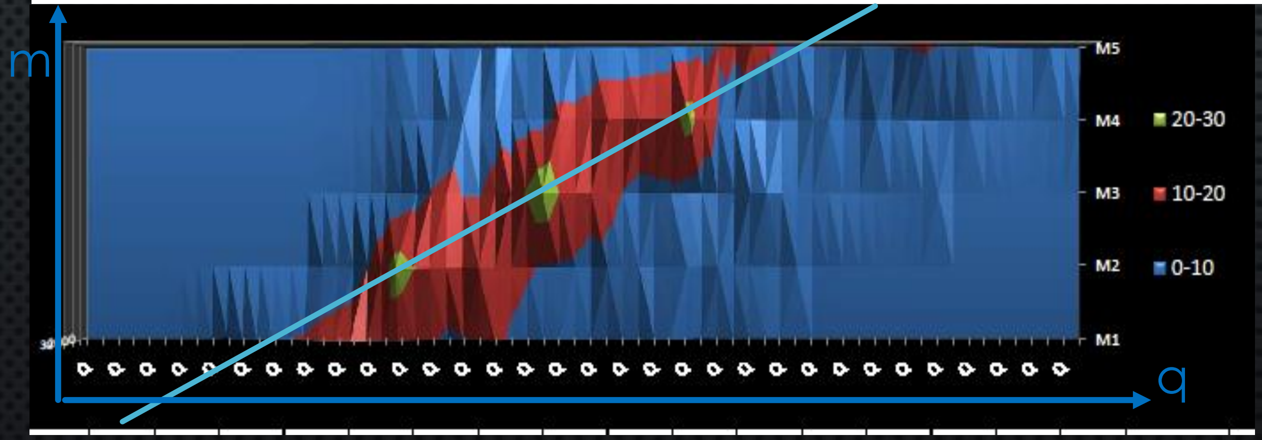



- The 5 tested *m* values displayed as a function of *q* → local maximum
- The three highest peaks are selecting lines crossing in the same point (vertex)
- In the (m,q) space 3 aligned points represent a vertex → vertex signature

# STRAIGHT LINE FIT EXAMPLE: Y=MX+Q



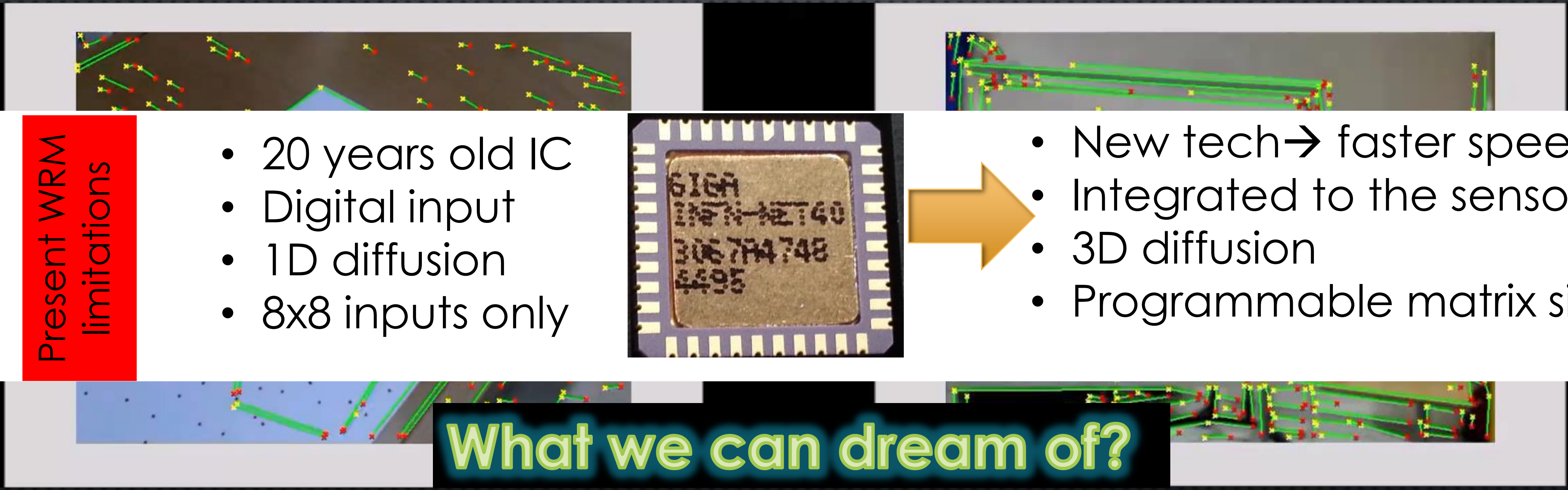



We can treat the parametric plane as new input data → in the new parametric pane m', q' of the line represents the vertex

# FROM PARTICLE TRACKS TO FAST IMAGE ANALYSIS

- THE IDEA: IN HEP WE CAN RECOGNIZE SIMPLE 1D TRAJECTORIES IN ~$10^{-9}$ S
- IN COMPUTER VISION WE NEED TO PROCESS 1 M PIXEL IN ~ $10^{-2}$ S
- WE CAN USE $10^{7}$ LONGER TIME TO PERFORM MUCH COMPLEX PATTERN RECOGNITION
- REBUILD THE IMAGE STRUCTURES IN TERMS OF A PARAMETRIC REPRESENTATION OF ITS ELEMENTS



**Present WRM limitations**

- 20 years old IC
- Digital input
- 1D diffusion
- 8x8 inputs only

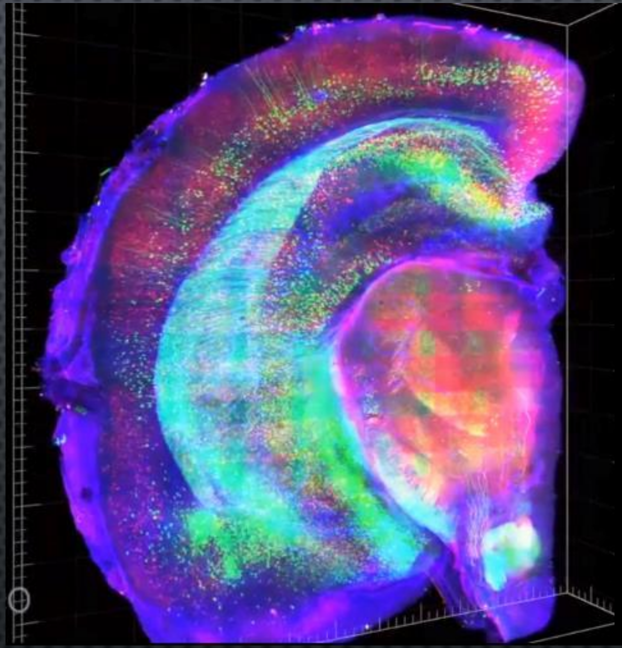
- New tech→ faster speed
- Integrated to the sensor
- 3D diffusion
- Programmable matrix size

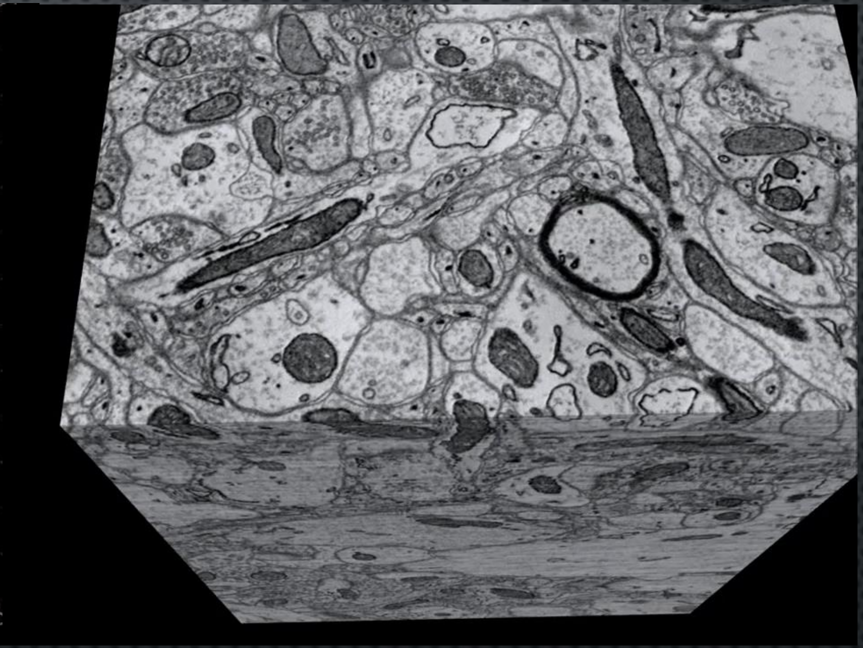
**What we can dream of?**

# FROM FRONTIER SCIENCE TO FRONTIER PROBLEMS

DREAM APPLICATIONS WITH A COMMON BOTTLENECK: UNTREATABLE ROW DATA THROUGHPUT

- REAL TIME PROCESSING FOR MEDICAL APPLICATIONS
- NEURON SYSTEMS READOUT

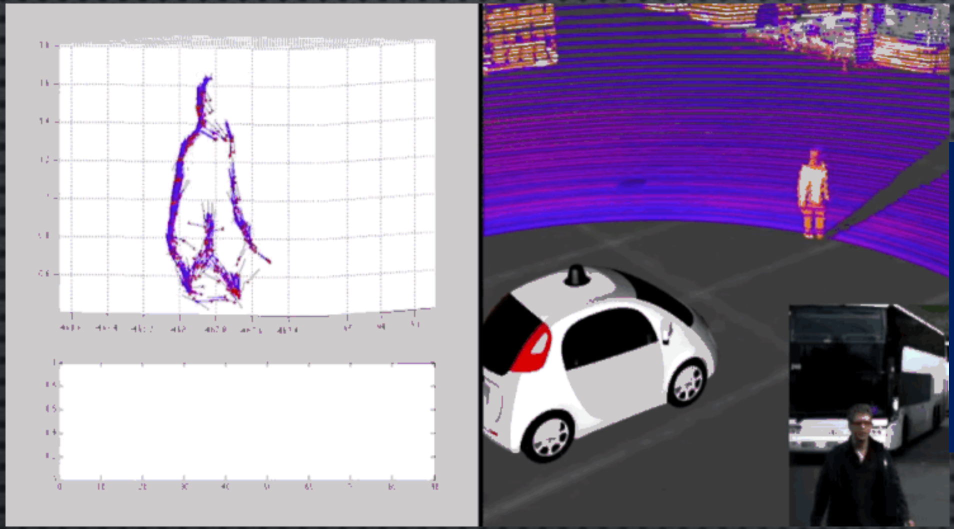




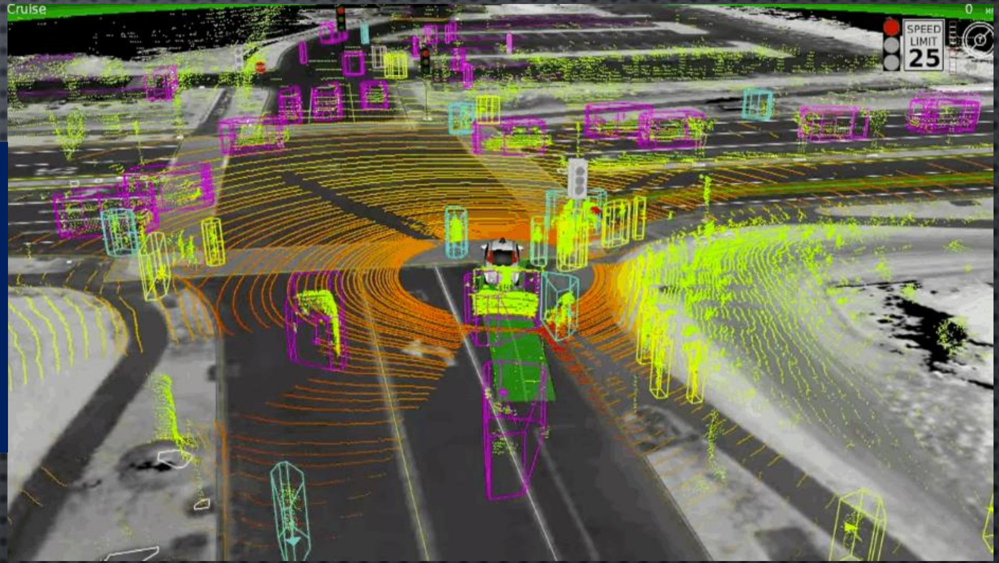
WORST THAN EVER: LOW POWER CONSTRAINTS FOR MOBILE APPLICATIONS

- AUTONOMOUS ROBOTS,
- SMART PROSTHESIS
- …AND SELF DRIVING CARS

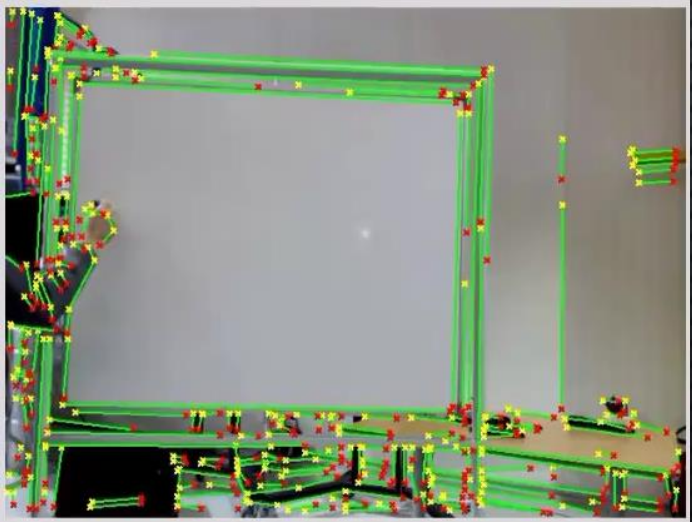
# FAST AND CAUTIOUS: CARS AND ROBOTS



How the Google Car sensors see the worls




How the a simple WRM sees objects through lines



Fast light and low power smart Front End is what would enable mobile smart devices challenge: low weight low power high performance.

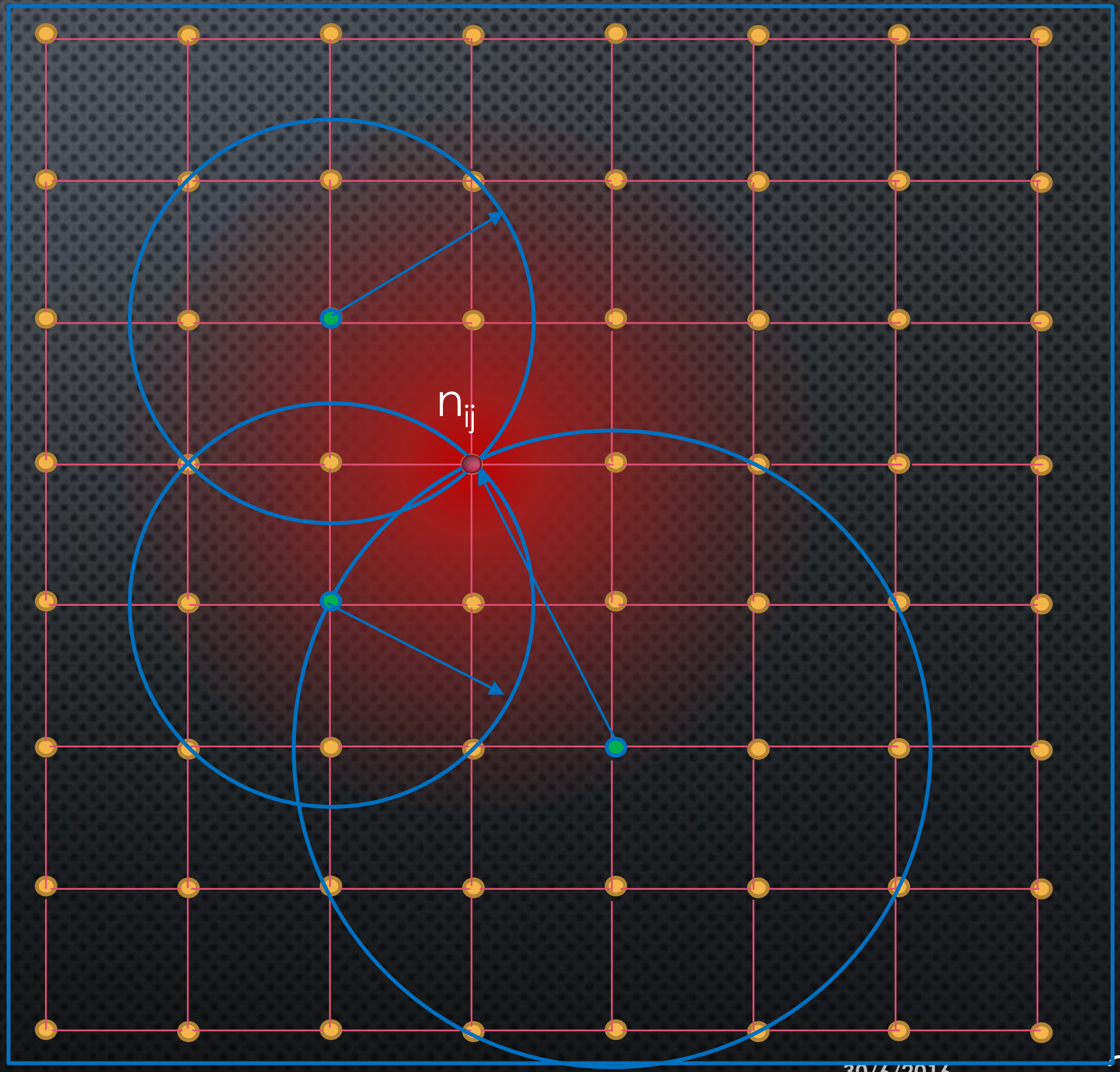
We propose to develop the next generation of WRM structures to face this challenge

# THANK YOU FOR THE ATTENTION

# BACKUP

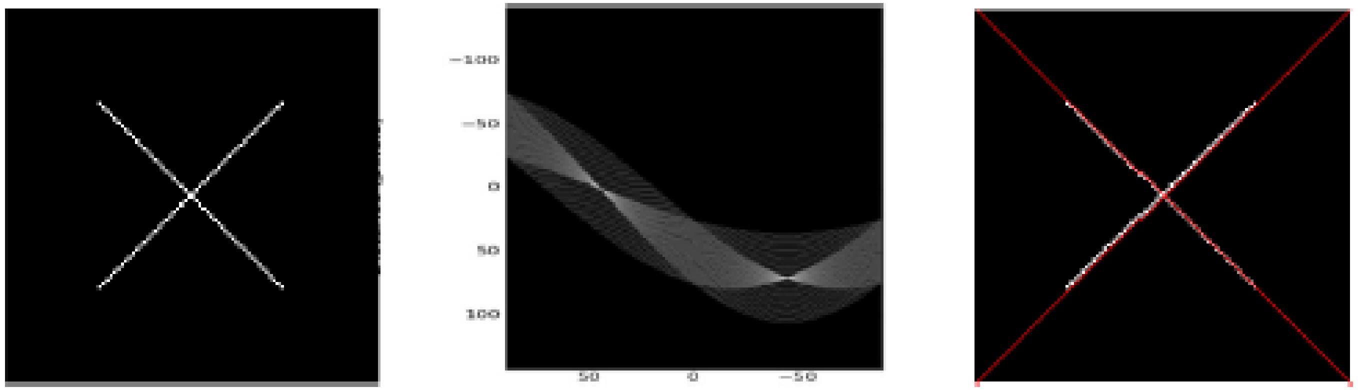
# WEIGHTING RESISTOR MATRIX (WRM) CONCEPT

- A GIVEN NODE $N_{IJ}$ IS SET TO A POTENTIAL V WHICH IS DIFFUSED TO THE NEIGHBORING NODES WITH A LAW:
  - $V_{KL}=(1/W)^{|N_{IJ}-N_{KL}|}$
  - W IS THE DIFFUSION CONSTANT WHICH DEPENDS ON THE RESISTIVE PARTITION
  - $V_{KL}$ IS CORRELATED TO THE DISTANCE
- BY MEASURING IN 3 POINT V IT IS POSSIBLE TO DETERMINE THE POSITION OF $N_{IJ}$
- BY MEANS OF AN APPROPRIATE SET OF MEASUREMENT IT IS POSSIBLE TO EXTRACT INFORMATION ON THE POINT POSITION
- **HOW TO EXTRACT GLOBAL INFORMATION ABOUT A PATTERN OF POINTS?**



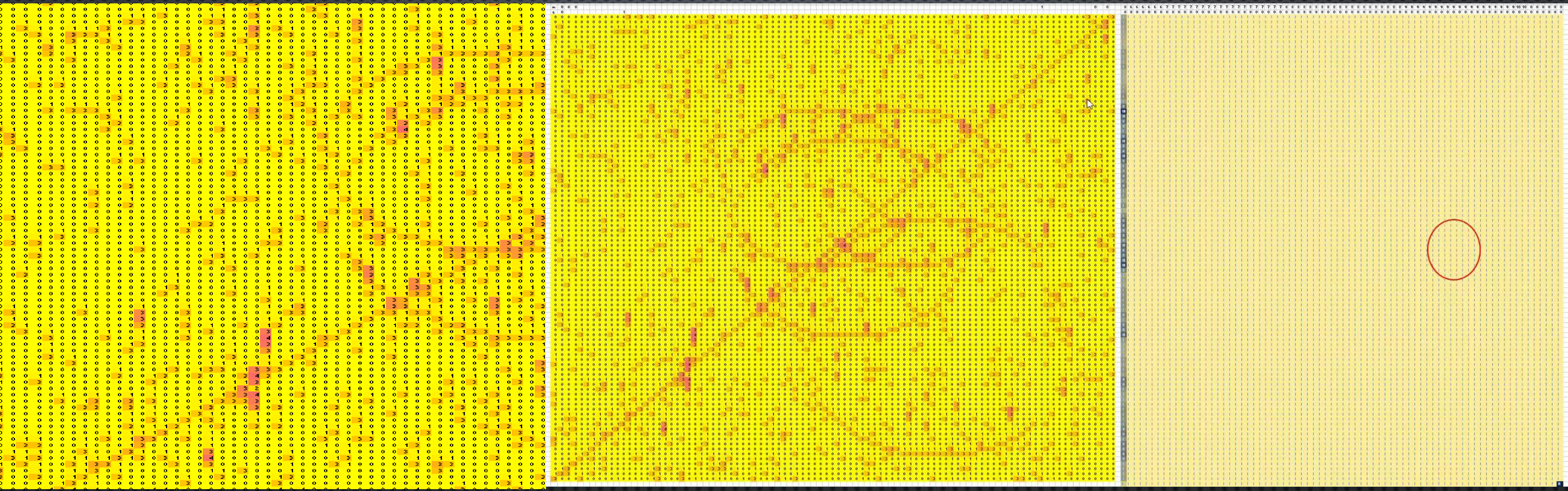
# Hough Transform

Describe straight lines in polor coordinates, $x cos(\theta) + y sin(\theta) = r$. For any "on" pixel $(x_0, y_0)$, compute a set of possible $(\theta_i, r_i)$.




(a) Input (b) Hough (c) Detected lines

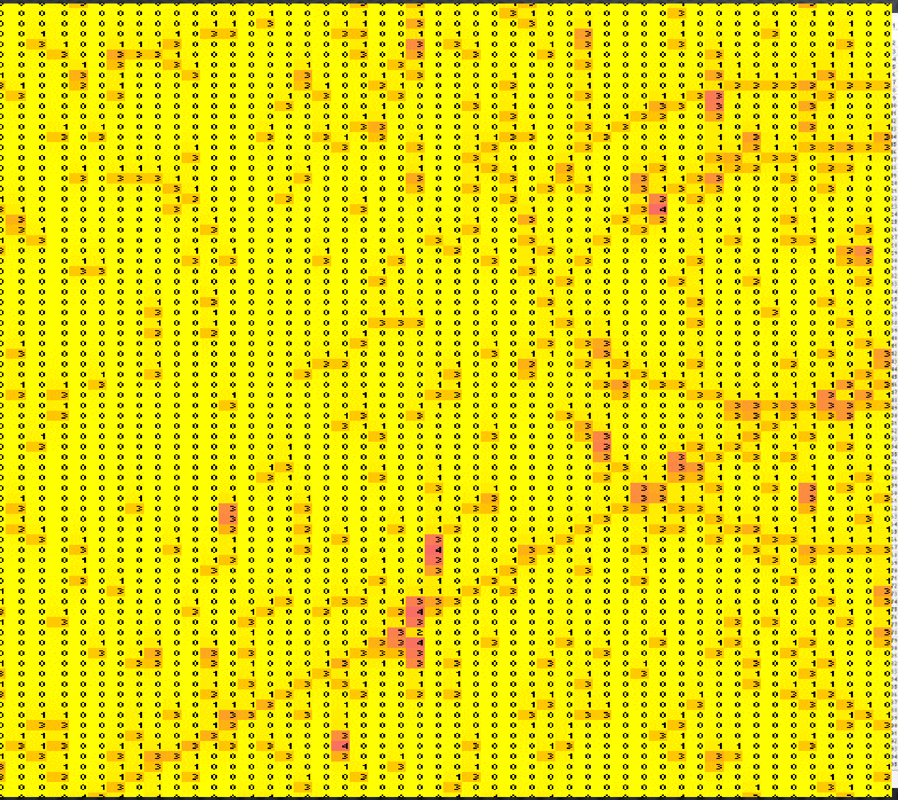
# SIMULATION OF 100X100 PIXEL MATRIX

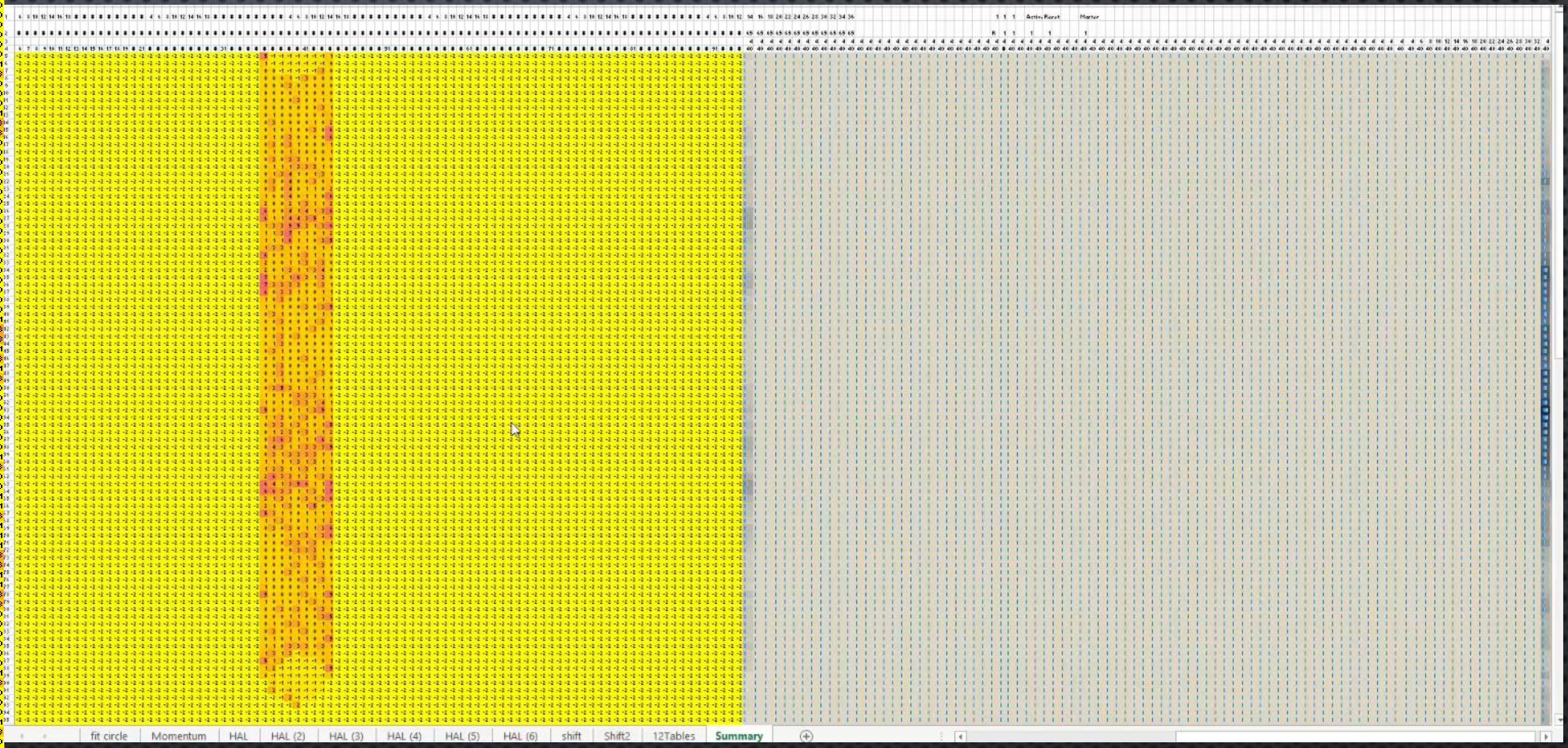


CRUCIAL: S/N increases with resolution → increase the number of degrees of freedom of the fit

- Shown how the data transform in the connection space
- *m* is looping {0…2} in steps of 0.1
- Each *m* test is a column inserted in the parametric space

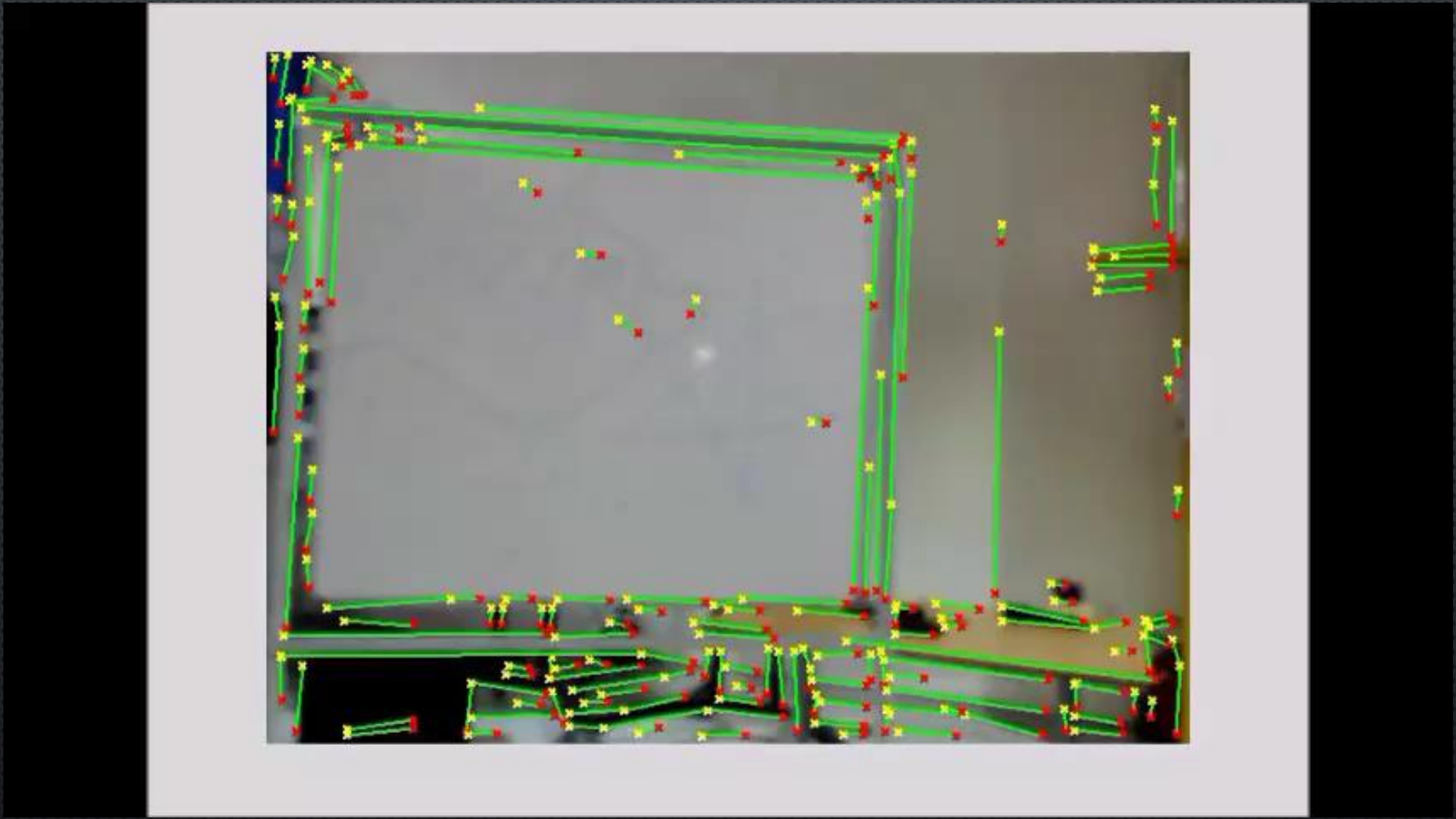
# SIMULATION OF 100X100 PIXEL MATRIX





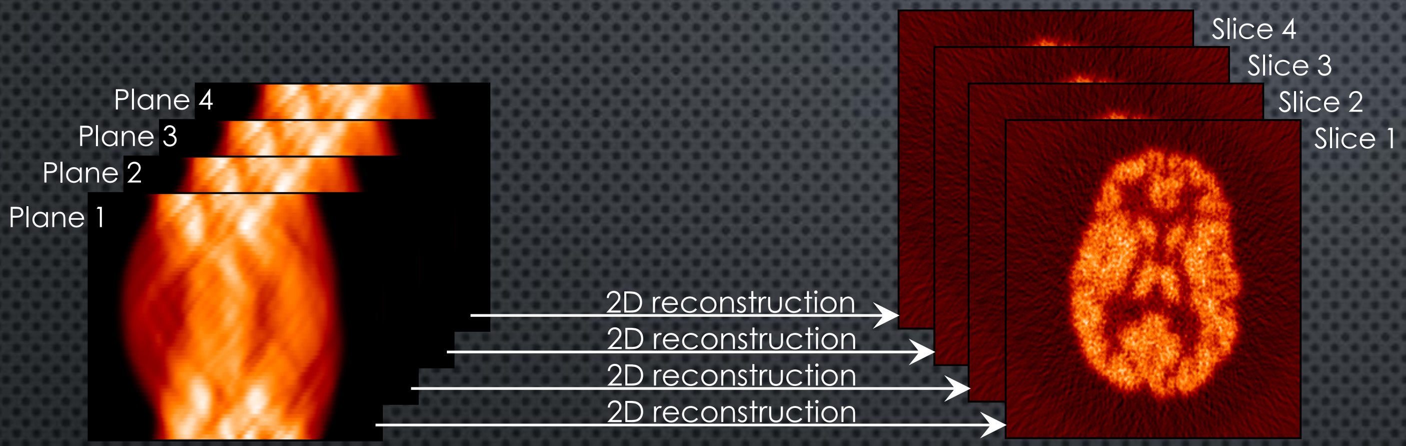

CRUCIAL: S/N increases with resolution → increase the number of degrees of freedom of the fit

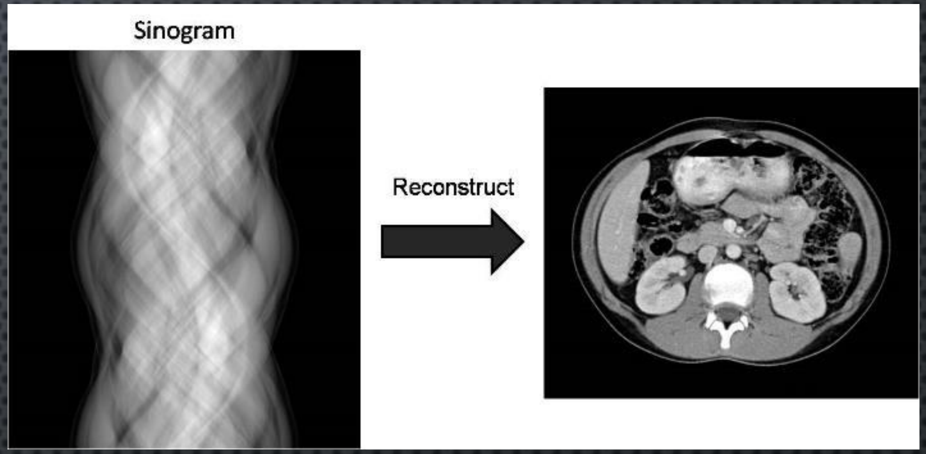
- Find the circle: loop on x, *R*
- 3D → need to do slices in parametric space
- All the steps needed? NO!!!!
- **Interpolation in parametric space**



# MEDICAL APPLICATIONS?

- DATA REDUCTION IN MEDICAL IMAGING → SMART FRONT END FOR PET/CT?



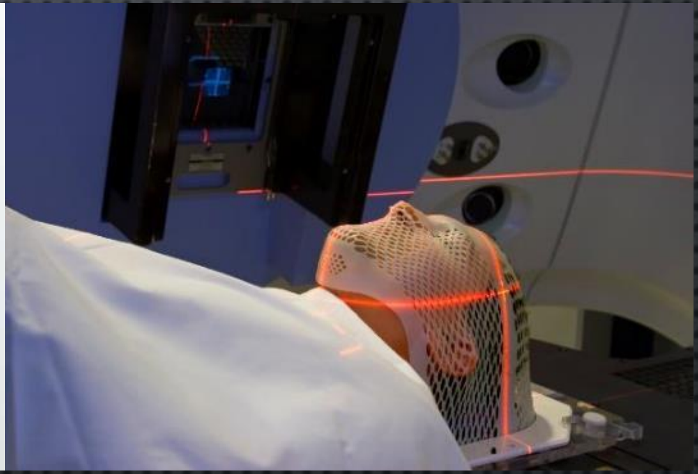




THE SINOGRAM IS VERY MUCH SIMILAR TO THE HOUGH TRANSFORM AND THE WRM IS ESSENTIALLY A VERY FAST HOUGH TRANSFORMER (IN ITS BASIC VERSION)

- FOR PET →COULD BE USEFUL FOR CORRECTING FOR SMALL MOVEMENTS OF THE PATIENT?

# MEDICAL APPLICATIONS?

- HADRONTHERAPY →COULD BE USEFUL FOR KEEPING THE BEAM ON A MOVING TARGET?

- FOR INSTANCE IT COULD BE POSSIBLE TO TRACK THE ORGAN IN REAL TIME WITH MM PRECISION  USING A FAST ARTIFICIAL VISION SYSTEM