“Unsupervised Learning and Real World Applications”

Machine IQ—Current Status of Computational Intelligence

Harold Szu, a Founder(INNS), Fellows (IEEE,OSA,SPIE,AIMBE), Academician (RAS)

Dir.Prof., Digital Media RF Lab

Dept. ECE, GWU

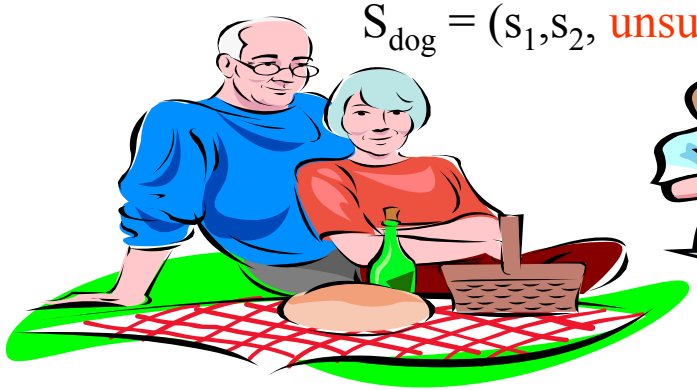
www.student.seas.gwu.edu/~dmlab

Unsupervised Redundancy Reduction to Features
Supervised learning is only for label of Features

Childhood learning experience of a "dog" begins with highly redundant inputs-
playful experience of eyes, ears, etc. 5 senses10 dimensional –
not by Fisher neighborhood separation

5-sense data $X_{dog}(t)$: ***effortlessly*** (fusion [W] X_{dog})→ $S_{dog}(t)$

$S_{dog} = (s_1, s_2,$ unsupervised features$)_{dog}$ fuzzy dog label by supervision



Not necessarily by separating dogs from Cats

Parents call out "a dog", a label of fuzzy linguistic nature
of which the vector span of the subspace is self discovered
by unsupervised manner, for no parents ever defined the dog.

Outlines:

1. New Direction for R/D

Brian-Center rather PC-Center Info Science & IT

e.g. 3rd Gen phone, Next Gen Internet Smart Sensors Web

2. Generalized Information Theory beyond Shannon

Entropy S Concept for real world open dynamic energy exchanging E equilibrium systems for power of pair sensing info input

e.g. in legacy electrical power-line.

3. Nano-Robot :a multiplexing atomic force microscope, e.g. by product: Biomimetic Imaging Fovea Camera by 1-D Carbon Nano Tube for its quantum mechanical low noise characteristics for early tumor Detection Infrared Spectrum Thermometer---Consumer Electronics Dual Camera replacing Hg

4. Theory & Mathematics of Neural Nets, Annealing

Outline:

(i) Two observations turn out to be the necessary & sufficient conditions for **learning without teacher**, **breakthrough of unsupervised ANNs Szu** 1997

(ii) Apply **physics to physiology to derive** **Brain Info Theory generalizing Shannon info theory**.

(1) Why "Power of 2"?

$$Vector \qquad \vec{X}(t) = [A(t)\,?]\,\vec{S}(t)\,?$$

$$2\ x\ 5 = 10 \quad D \quad redundancy \qquad reduction$$

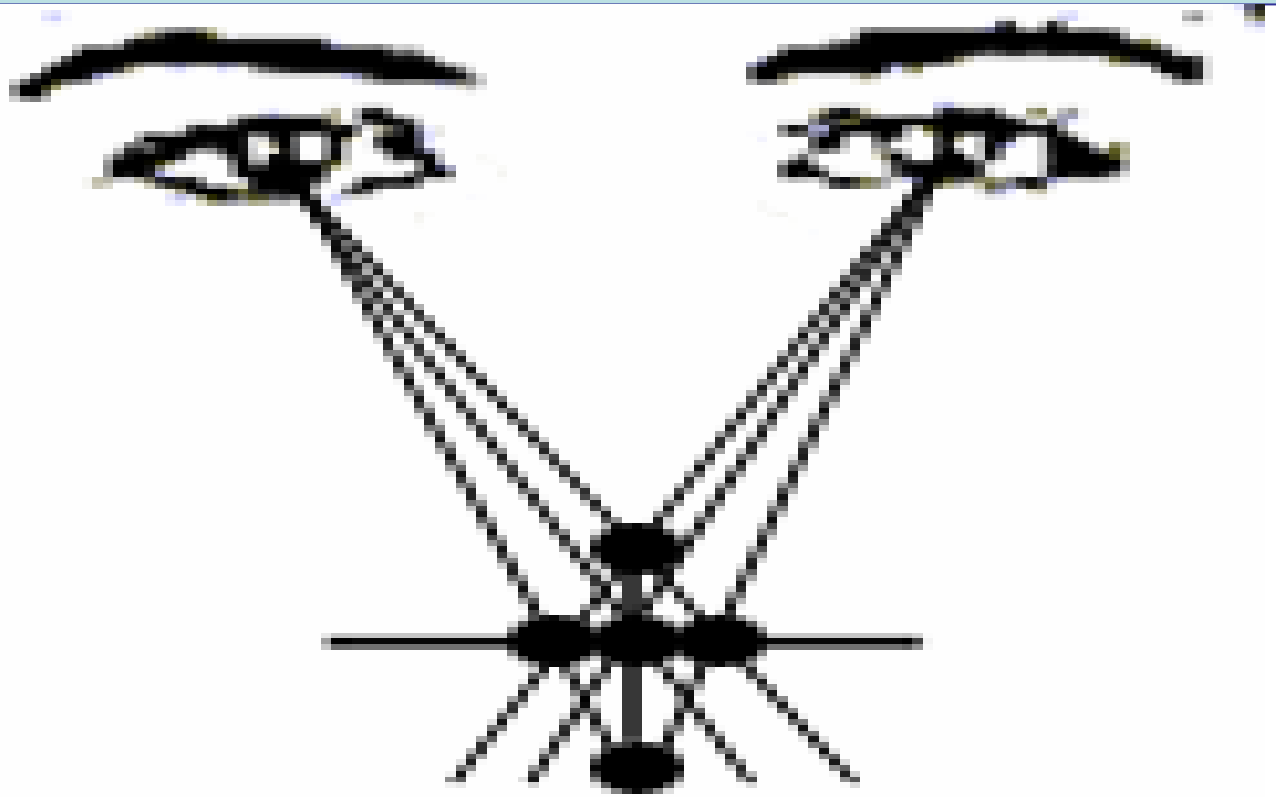
(2) Why "37°C" ?

But the output has one unique sensor!

"Homeostasis Learning" Szu, 1999 Rus. Acad. Sci.

Harold Szu, GWU Washington DC

Why do we have two eyes?
Why receptors input prefer pairs, while emitters are singles?



***Answers*:**

(1)Hardware fault tolerance;

(2)Stereovision—no range info David Marr Binocular Paradox

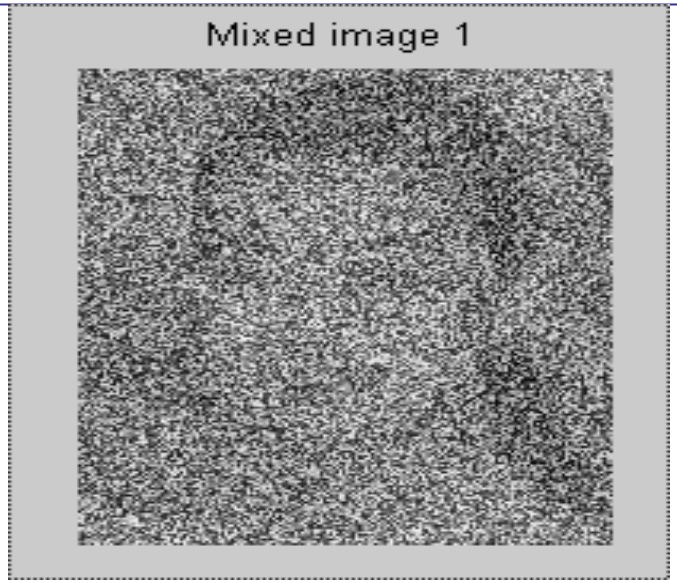
(3)Unsupervised learning by Coincidence Account

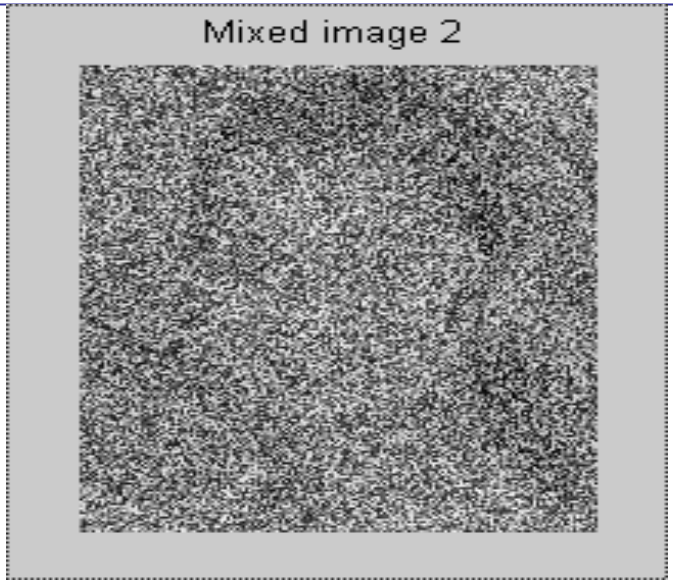
(4) All above (correct answer)

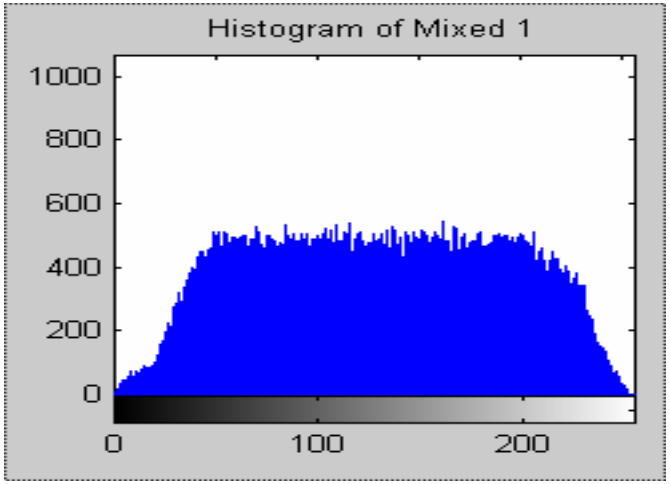
Blind Sources Separation gives early detection of life-death decision

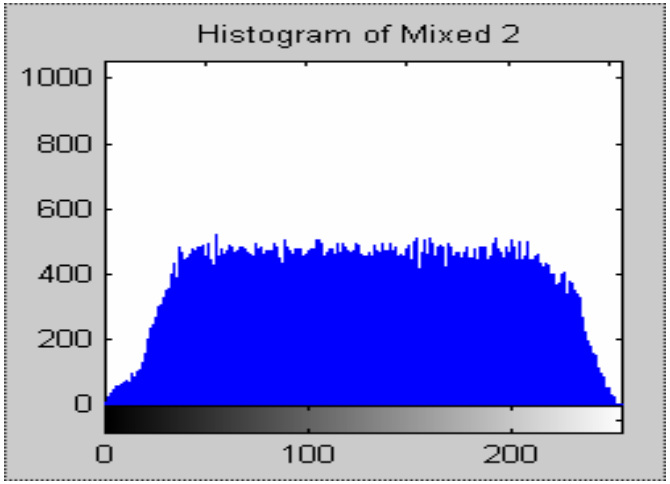
Two eyes could see through the fog---

two-eye a perfect restoration as opposed to one eye image processing is merely re-shoveling snow!

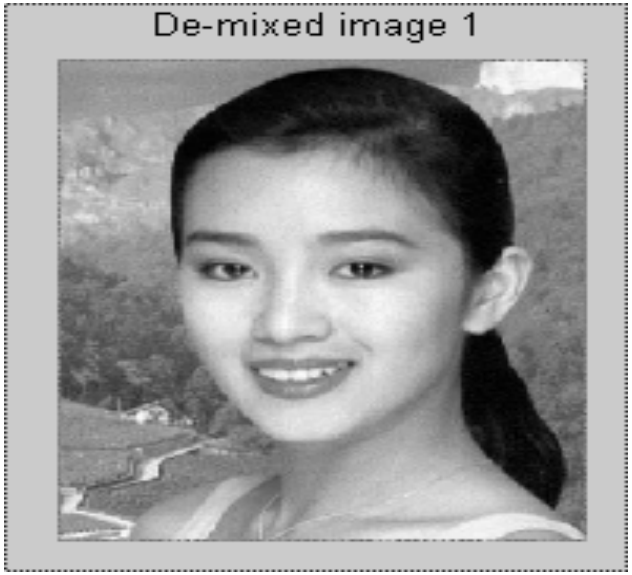


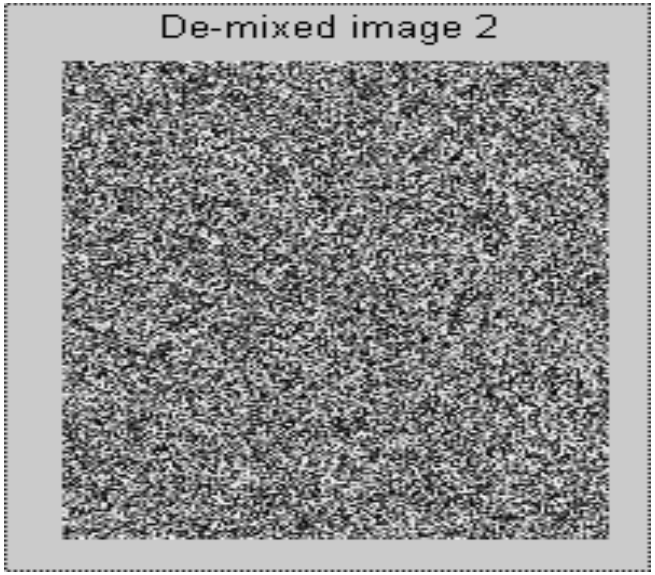


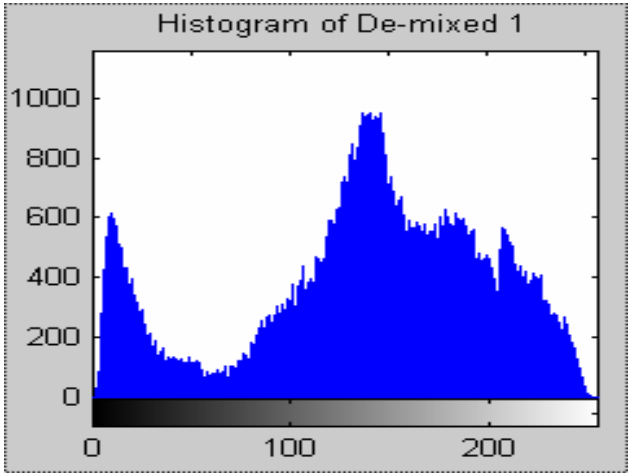




While agreements must be signals, disagreements, noises, that are universal and need no teacher!!







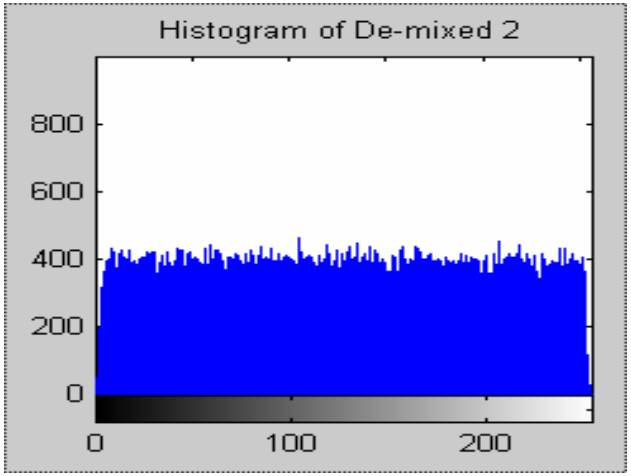
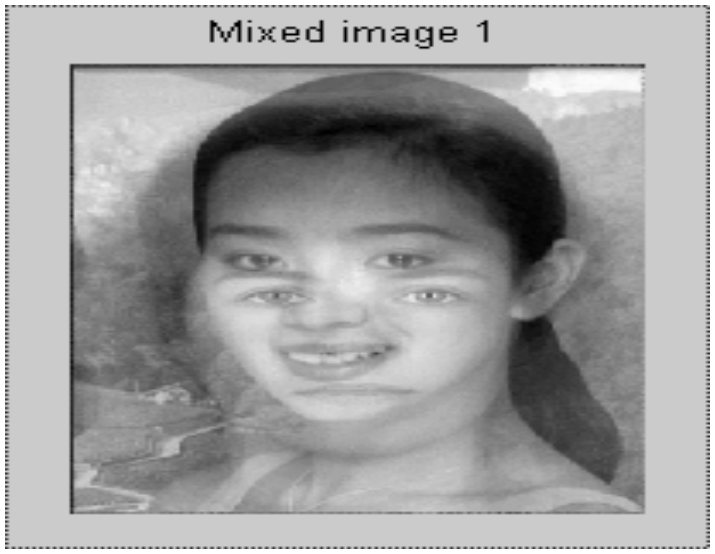
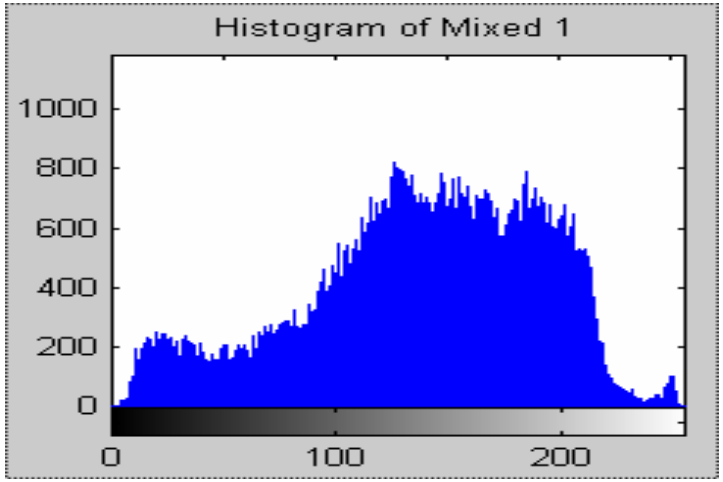
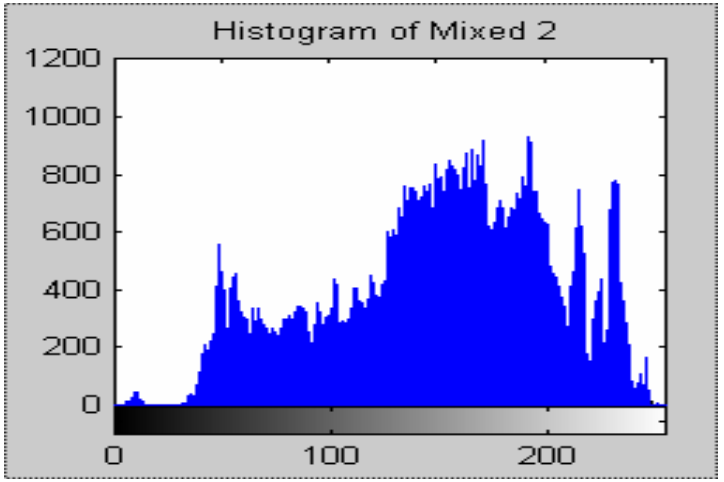


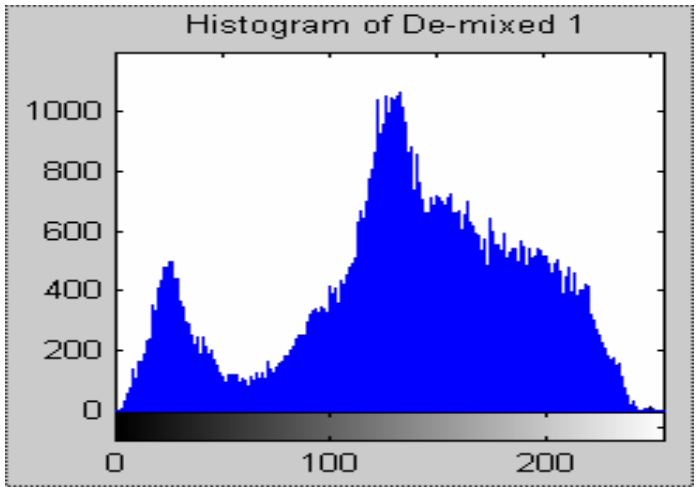
Illustration of Blind Source Separation

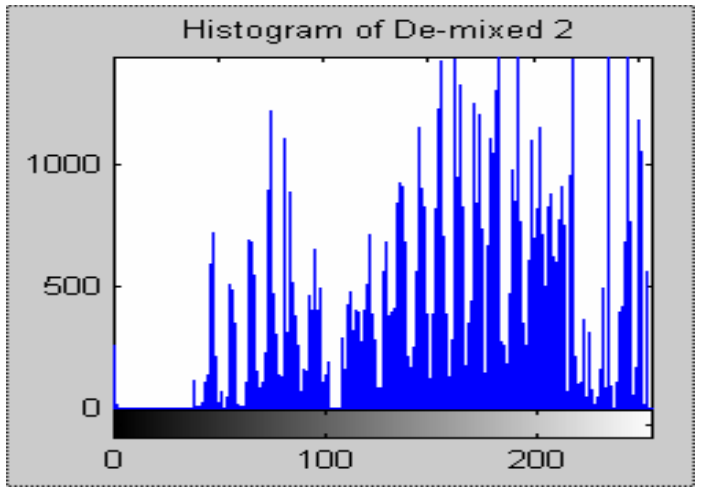






De-mixed images and histograms



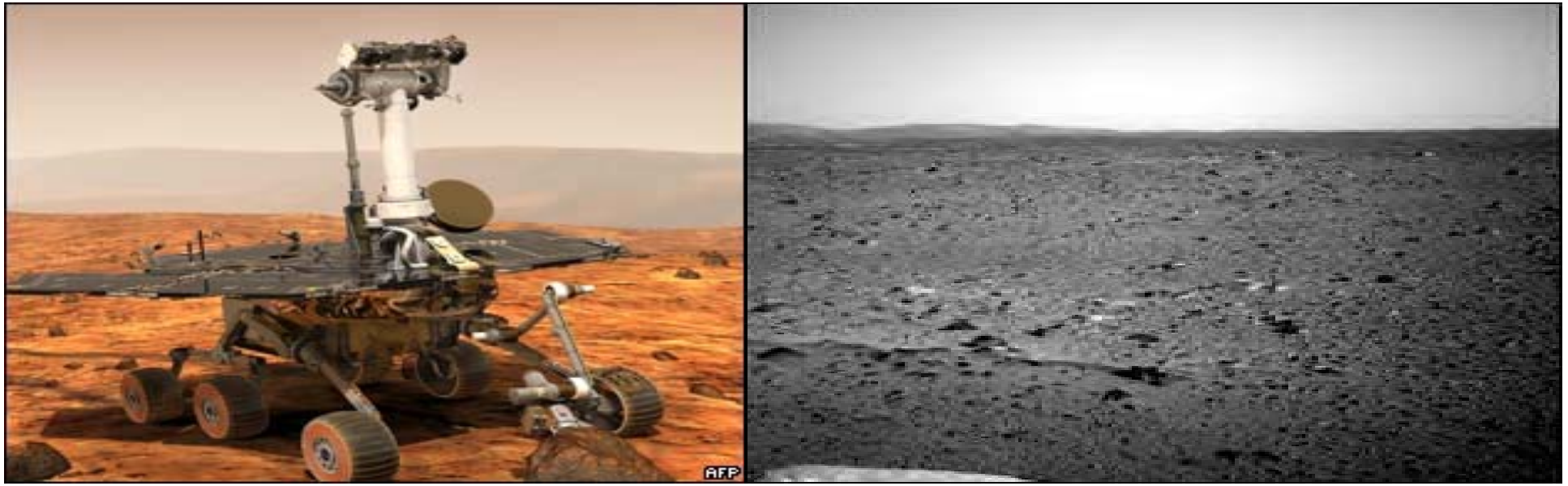


Real World Applications from Tank to Tumor need unsupervised MIQ/CI capability

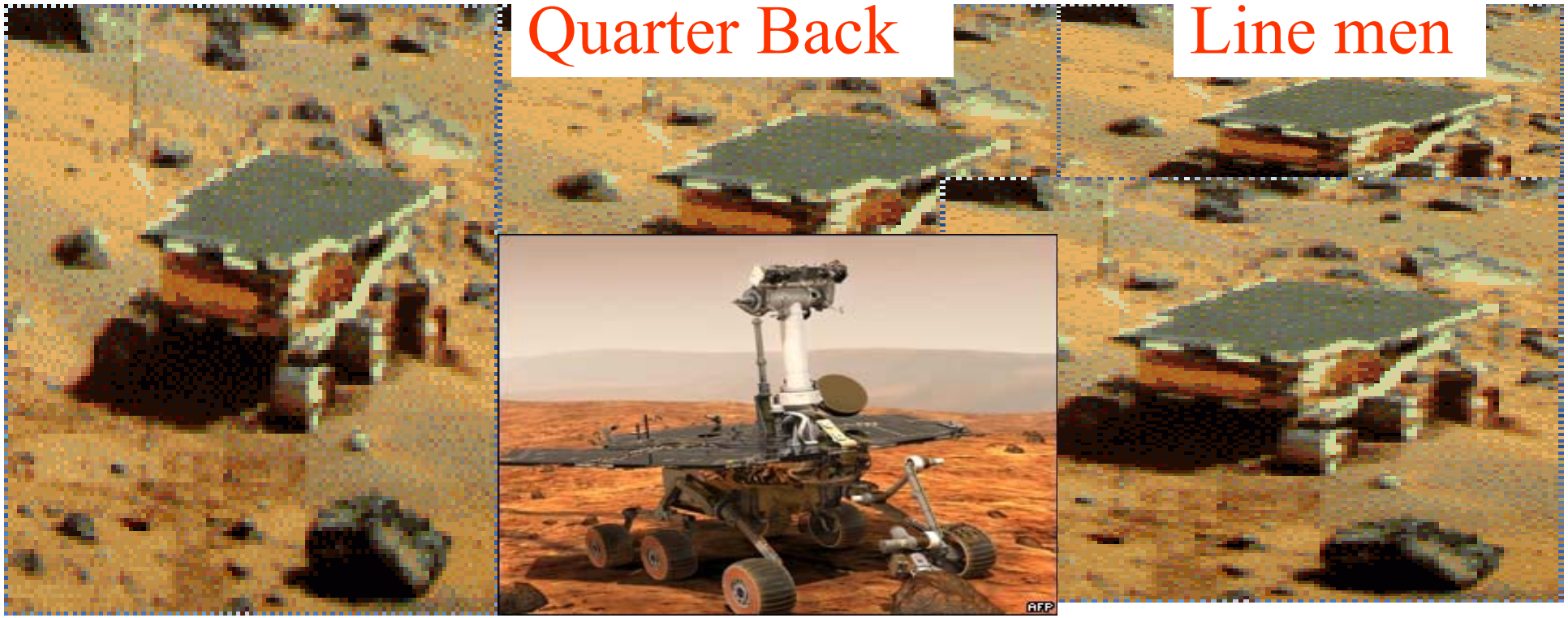
- NASA remote sensing in outer space without ground truth
- Mission To Moon 2010, To Mars !
- mini-UAV for Future Naval Capability e.g. Protect $2B Aircraft from $1M Cruise Missiles attack
- mini-UAV 3D Synthetic Aperture Radar seeing through forest, rainy foggy weather, sand storm
- NIH/NCI; NIH/NIBIB; molecular tagged imaging, early tumor detection

Motivation: futuristic robots need Swarm Team Partially Unsupervised Intelligence within their uncharted terrain

•Communication delay about 5 min. between Mission Control and Robots on Mars.
•Make any un-expected decision; when a single inexpensive robot has difficulty to deal with real world open set, a team of simple robots is considered.
•What can be expected from the swarm intelligence of Ants, Bees, Birds, Fishes, of which the basis of a min. sensor-communication?

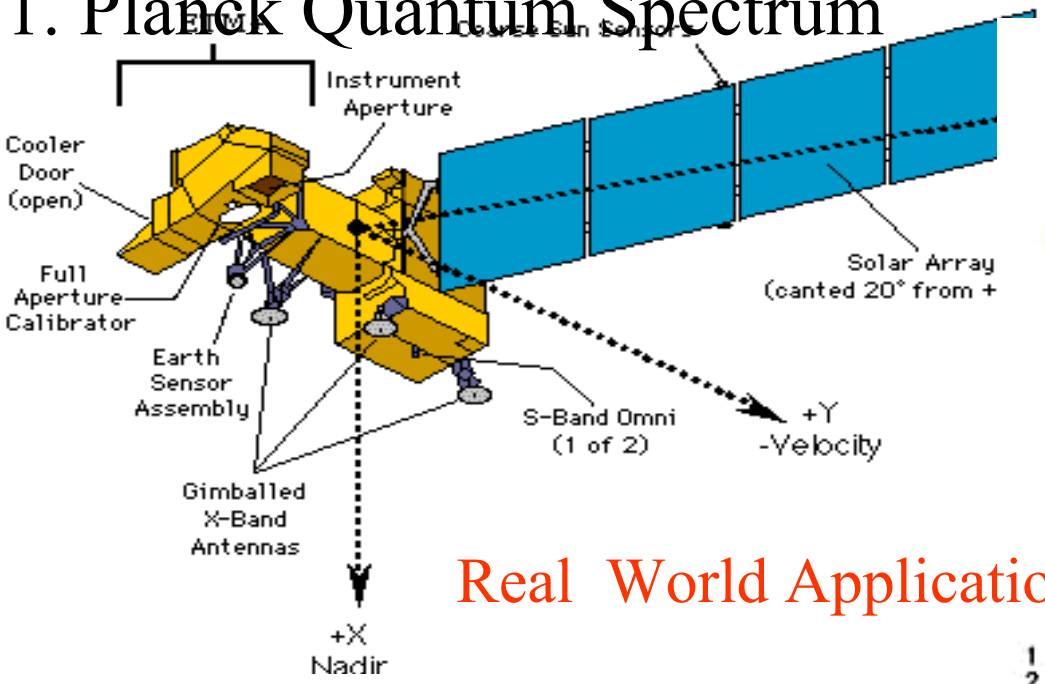


Martian UXVs needs Unsupervised Natural Intelligence

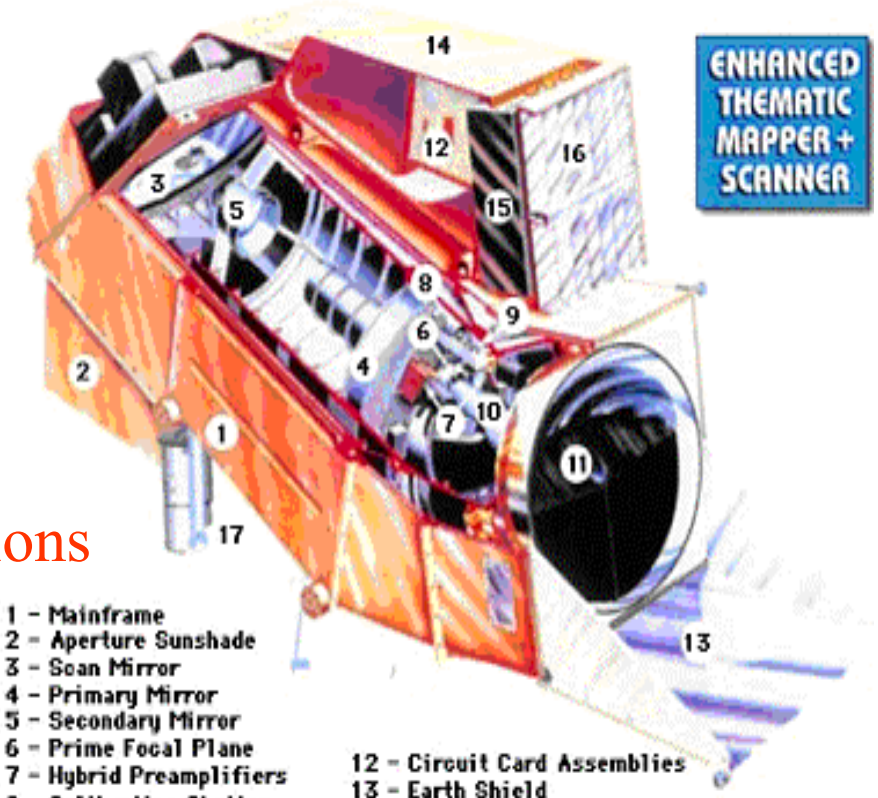


Martian pathfinders require a local adaptive and intelligent control because of about 5 min. time-delay from the ground station. Instead of one-to-one, we have proposed a quarterback robot (like a queen bee with 50%Machine IQ) controlling linemen robots (not unlike a queen bee controlling a hundred working bees)

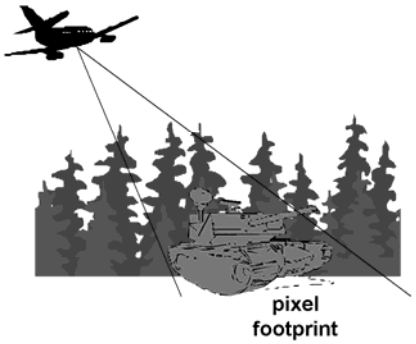
1. Planck Quantum Spectrum

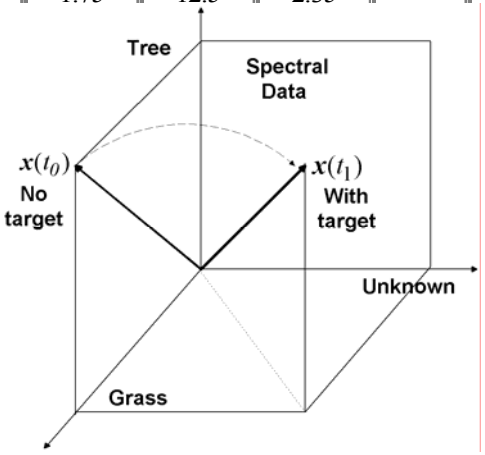


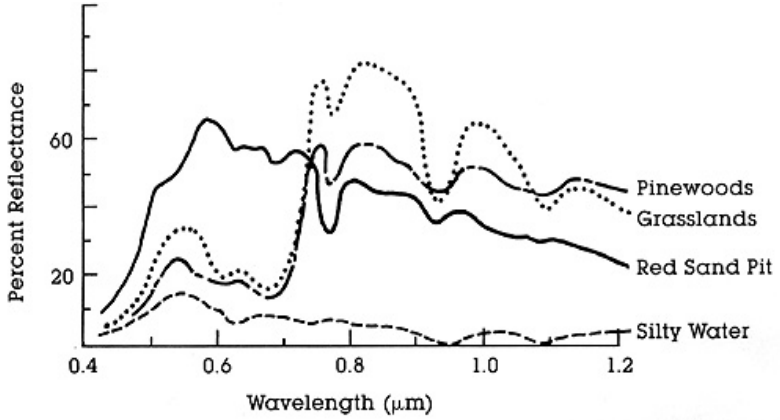
Real World Applications



TM AND ETM+ SPECTRAL BANDWIDTHS								
Bandwidth (μ) Full Width - Half Maximum								
Sensor	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7	Band 8
TM	0.45 - 0.52	0.52 - 0.60	0.63 - 0.69	0.76 - 0.90	1.55 - 1.75	10.4 - 12.5	2.08 - 2.35	N/A
ETM+	0.45 - 0.52	0.53 - 0.61	0.63 - 0.69	0.78 - 0.90	1.55 - 1.75	10.4 - 12.5	2.09 - 2.35	.52 - .90

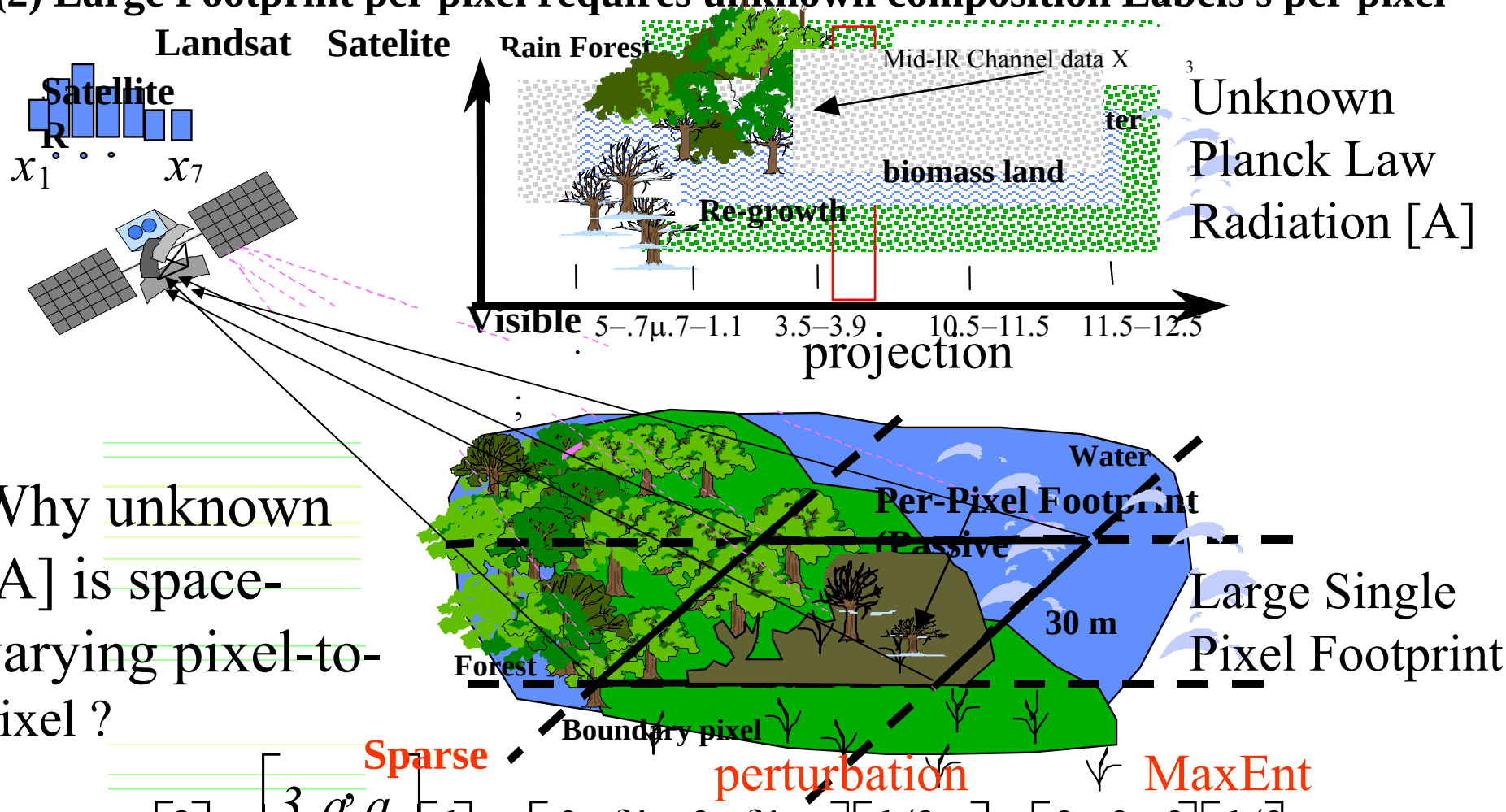






(1) Diurnal & seasonal variations yield unknown objects-spectral table matrix [A]

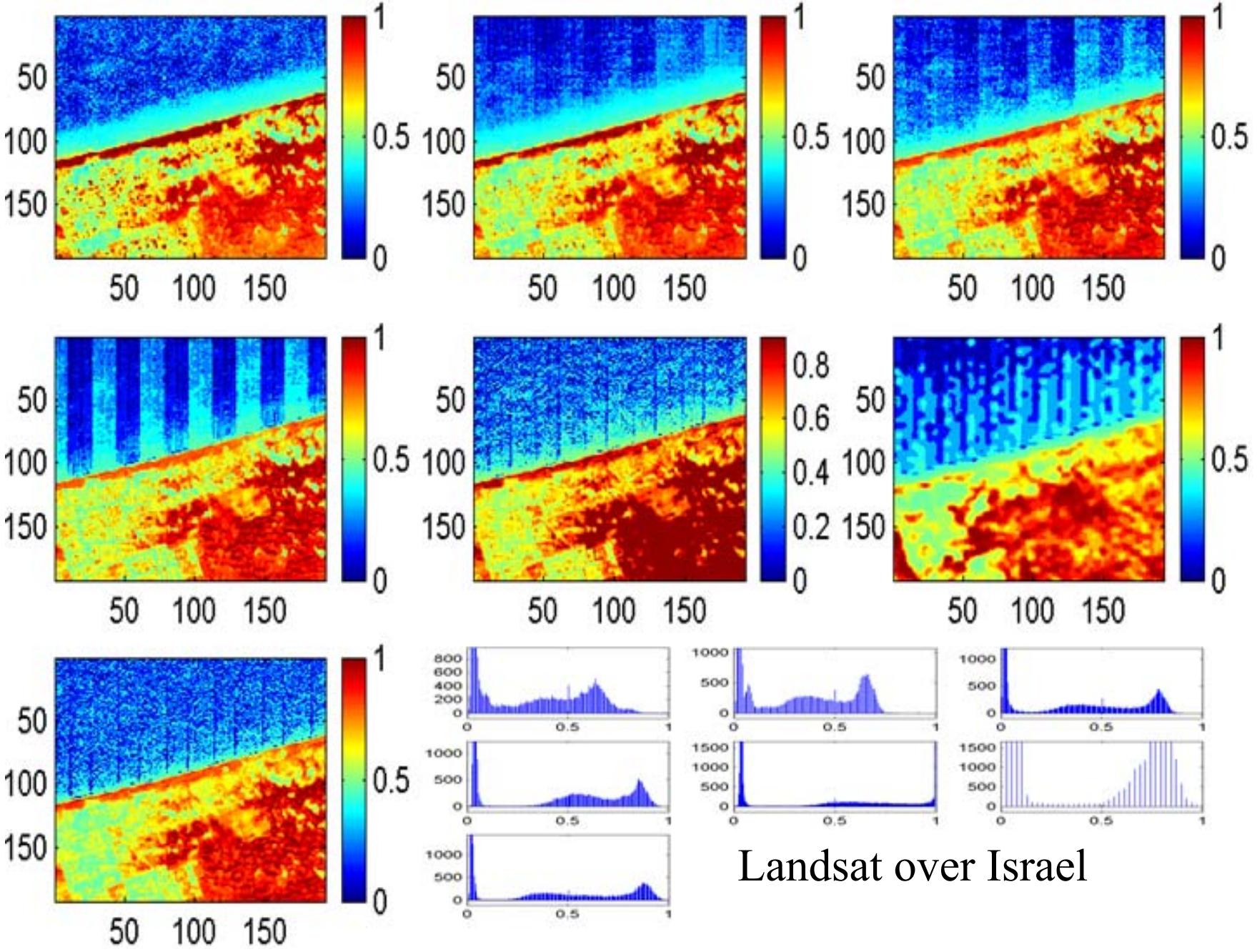
(2) Large Footprint per pixel requires unknown composition Labels s per pixel



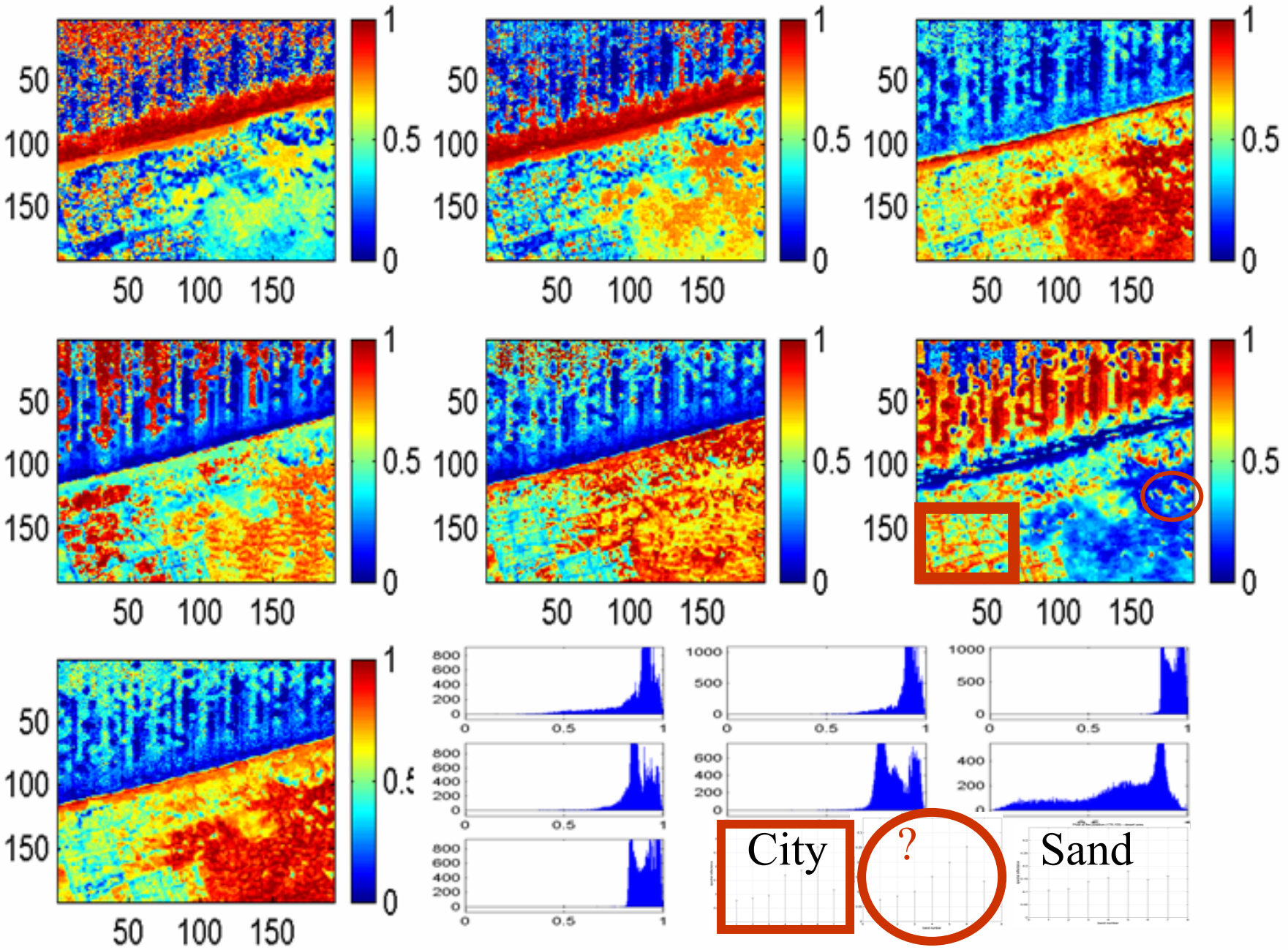
Why unknown [A] is space-varying pixel-to-pixel ?

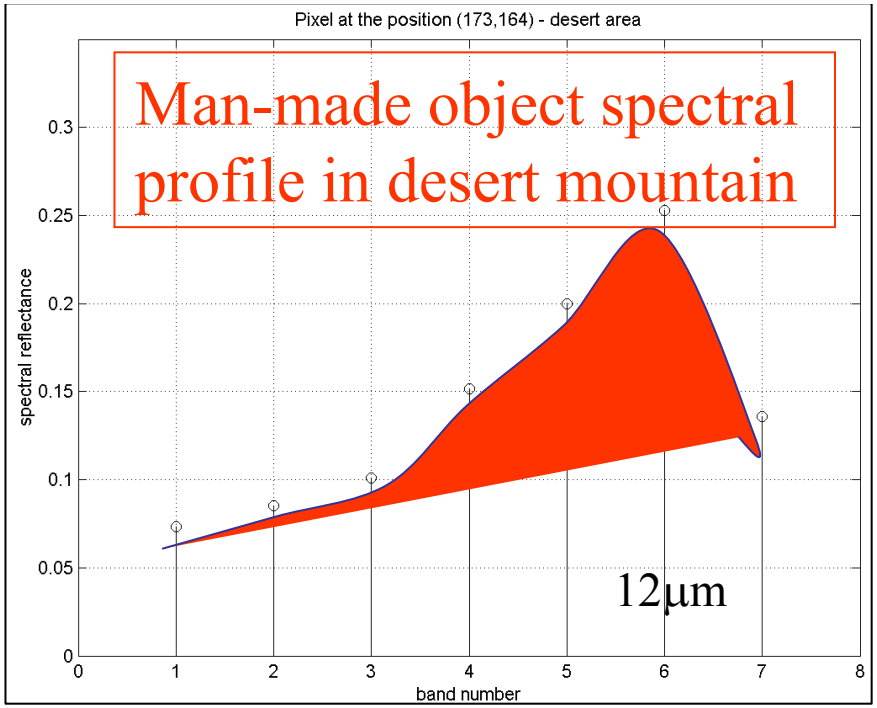
Sparse perturbation MaxEnt

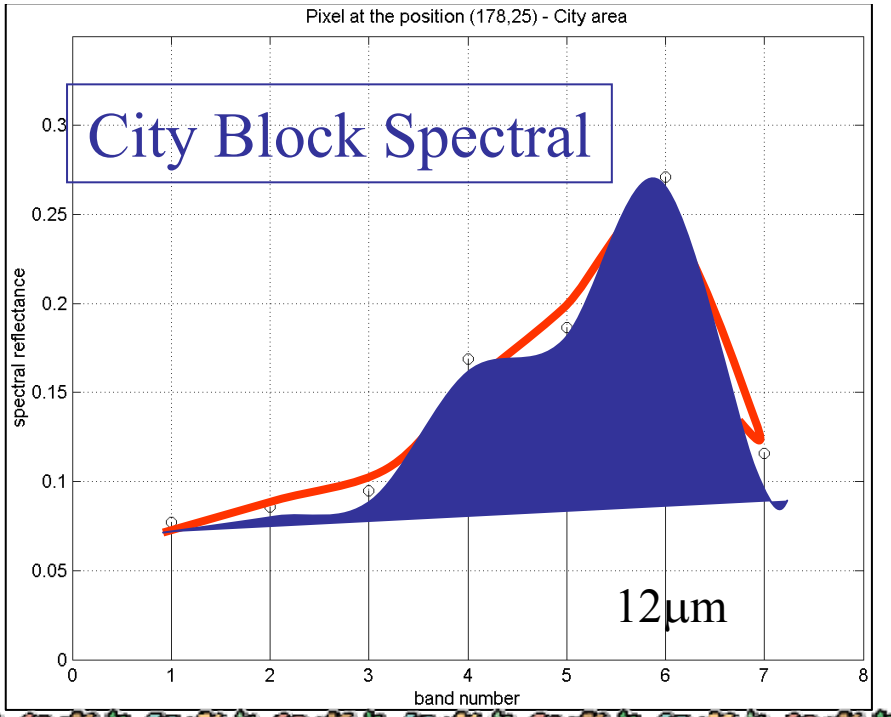
$$\mathbf{x} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 & a & a \\ 2 & a & a \\ 1 & a & a \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cong \begin{bmatrix} 3+\delta' & 3-\delta' & a \\ 2+\delta'' & 2-\delta'' & a \\ 1+\delta''' & 1-\delta''' & a \end{bmatrix} \begin{bmatrix} 1/2+\varepsilon \\ 1/2-\varepsilon \\ 0 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 3 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

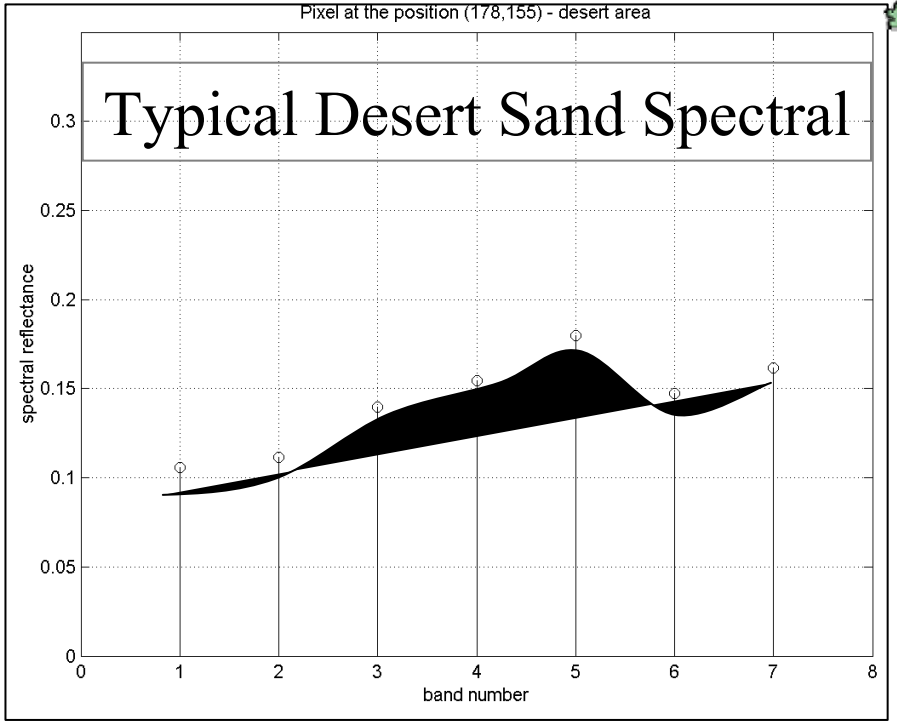


Landsat over Israel


50
100
150
1
0.5
0
City
?
Sand







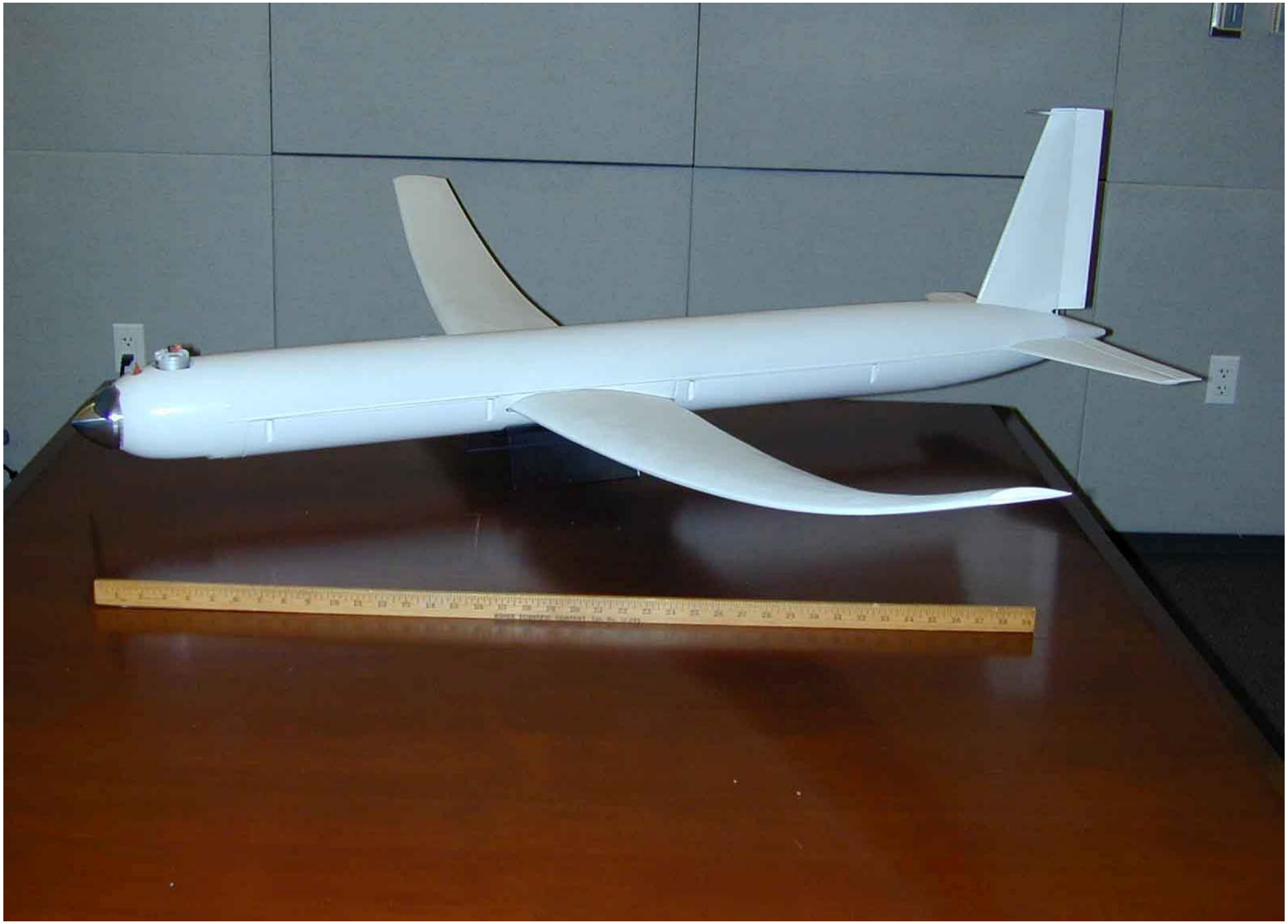
In Duda-Hart sense of unlabelled Data X LCNN discovers Planck Radiation Law as overlapping Fuzzy Memberships

Top left pixel (173,164): LCNN discovered a "man-made construction" in desert mountain area because of the correspondence pixel (178, 25) in City block shown Top right; whereas a "typical" pixel (178,155) in desert shown in Bottom left.

Portable Multispectral Video might prevent USS Cole terrorists attack

mini-UAVs provide covert Over-the-Hill/Horizon early warning

- Submarine pre-surfacing look out post
- Chem-bio-Nuclear fly in sensor collection
- Communication linkage ad hoc network
- Passive Synthetic Aperture Radar
- Passive EO/IR video surveillance

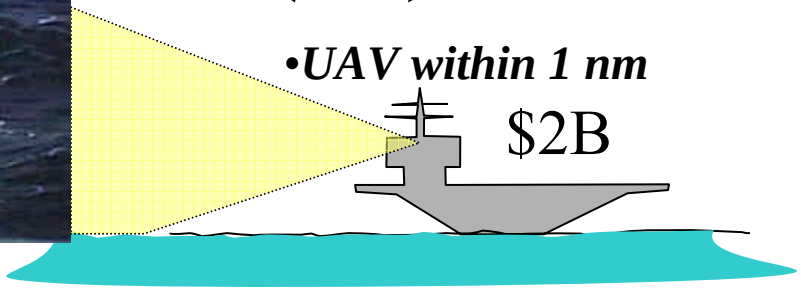


How to increase Stand-off Distance of Missile Defense?

In-situ mini-UAV can give OTH carrier's scouting needs, but the payload can not use cryogenic multispectral imaging.

Coverage:

- ***360-degrees-horizon***
- ***48 kilometers Range (26nm)***
- ***UAV within 1 nm***



- UAV can steer itself by means of Long IR FPA to point at incoming cruise missile plume looking for unique feature at Mach cone turbulent mixing range dependence of CO-CO_2 Mid-IR spectral lines.
- Thus, FNC platform protection can increase the standoff distance by unsupervised fusion in terms of two color IR by UAV as X-band Radar can not yet see OTH missile due to the earth curvature ocean waves.

Brain-Center informatics, rather Computer-Center IT needs unsupervised thus unbiased data/text mining

- We review supervised data mining such as Self Organization feature Map
- Expert knowledge is needed for bioinformatics
- Thus, instead of bioinfo we approach bioinformatics from macro-system of human society such as the journalism 6 W representation-who, where, when, how, what & why.

Two Breakthroughs of unlabelled mixture X=[A]S
Unsupervised Learning for identically unknown [A], called Independent Component Analysis (ICA)
j-pdf Factorization *r(x1,x2,x3,..) = r1(x'1) r2(x'2).*
Where X'=[W]X by Bell&Sejnowski of Salk. Amari of RIKEN, Oja of Finland, etc. 4 bi-annual conferences since 1997.
Another fundamental physics thermodynamic approach led by Szu
Fro real world space-variant mixing [A] in remote & tumor
To be Constrained by the isothermal equilibrium min. H=E-TS for Pixel-by-pixel Blind Source Separation (BSS)

Why do our brains have pairs of sensors kept at a constant temperature ?

2. Generalized Shannon Info Theory to Brain Info Theory

by minimum of Helmholtz thermodynamic free energy,

Min. $H = E - T_o S$ (Max. by Shannon under T_o=37°C)

Szu assumes analytic I/O $E=\mu([W]X - S) + \lambda([W]X - S)^2$ +Taylor

1997 Theorem for unsupervised sensory fusion: If info energy E is analytic in info I/O , (data X & feature S), then the necessary & sufficient conditions of unsupervised learning in Dude-Hart unlabelled data classifier sense are:(1) An intelligent brain is kept at constant temperature, e.g. human 37°C

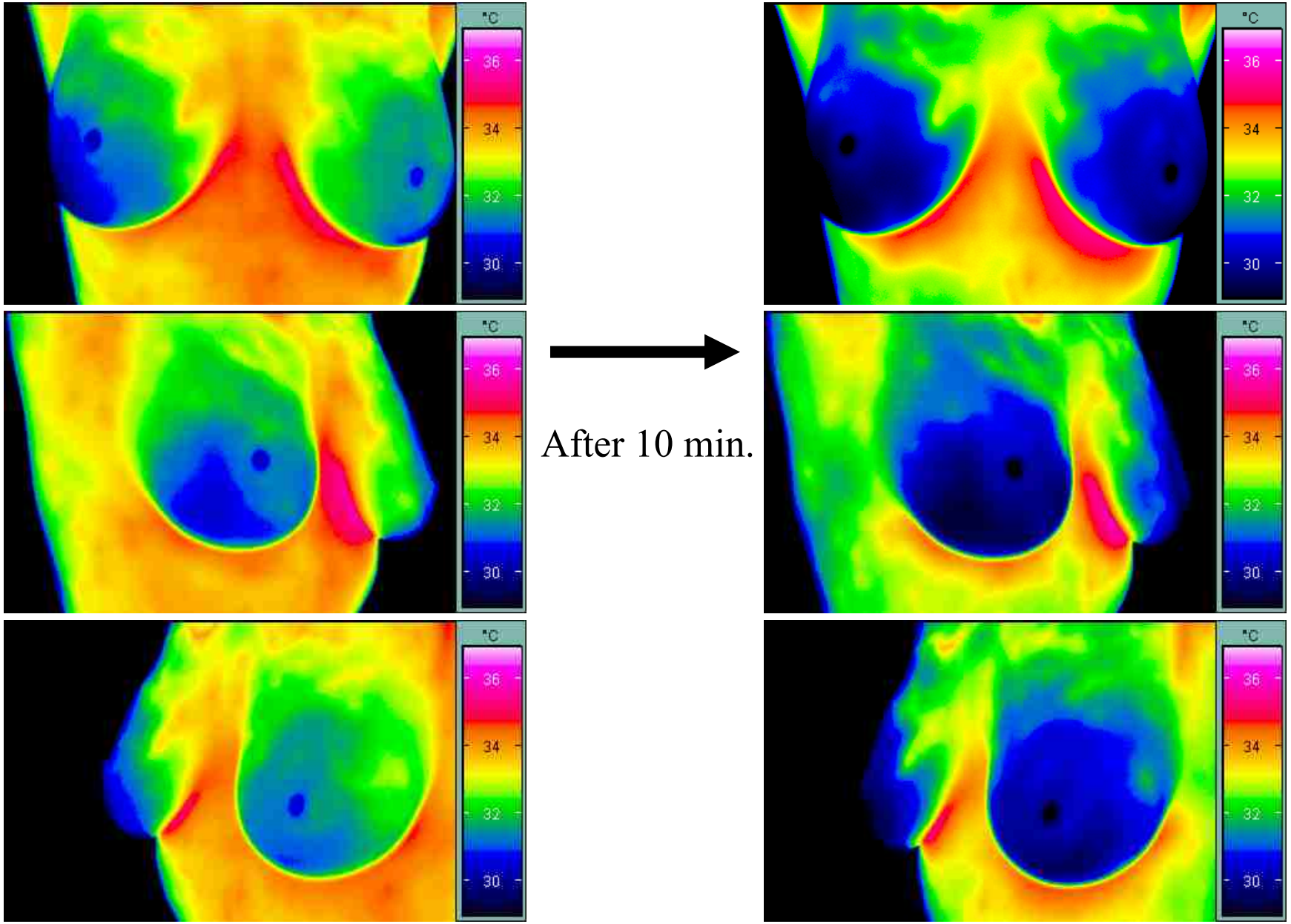
Unsupervised Learning

(2) All input sensors are Smart Pairs : "Power of Pairs In, Garbage Out"

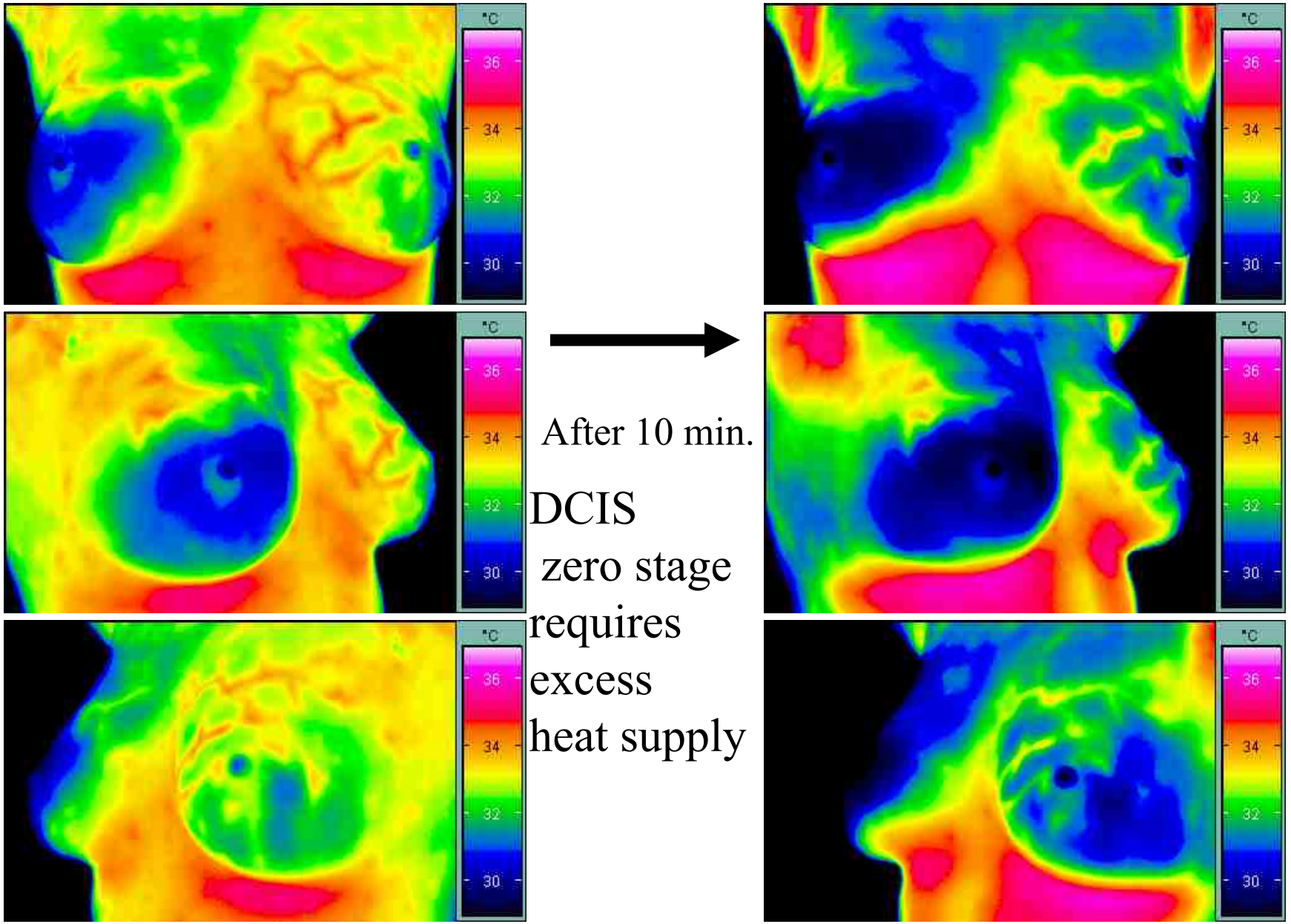
Information is kept within memory

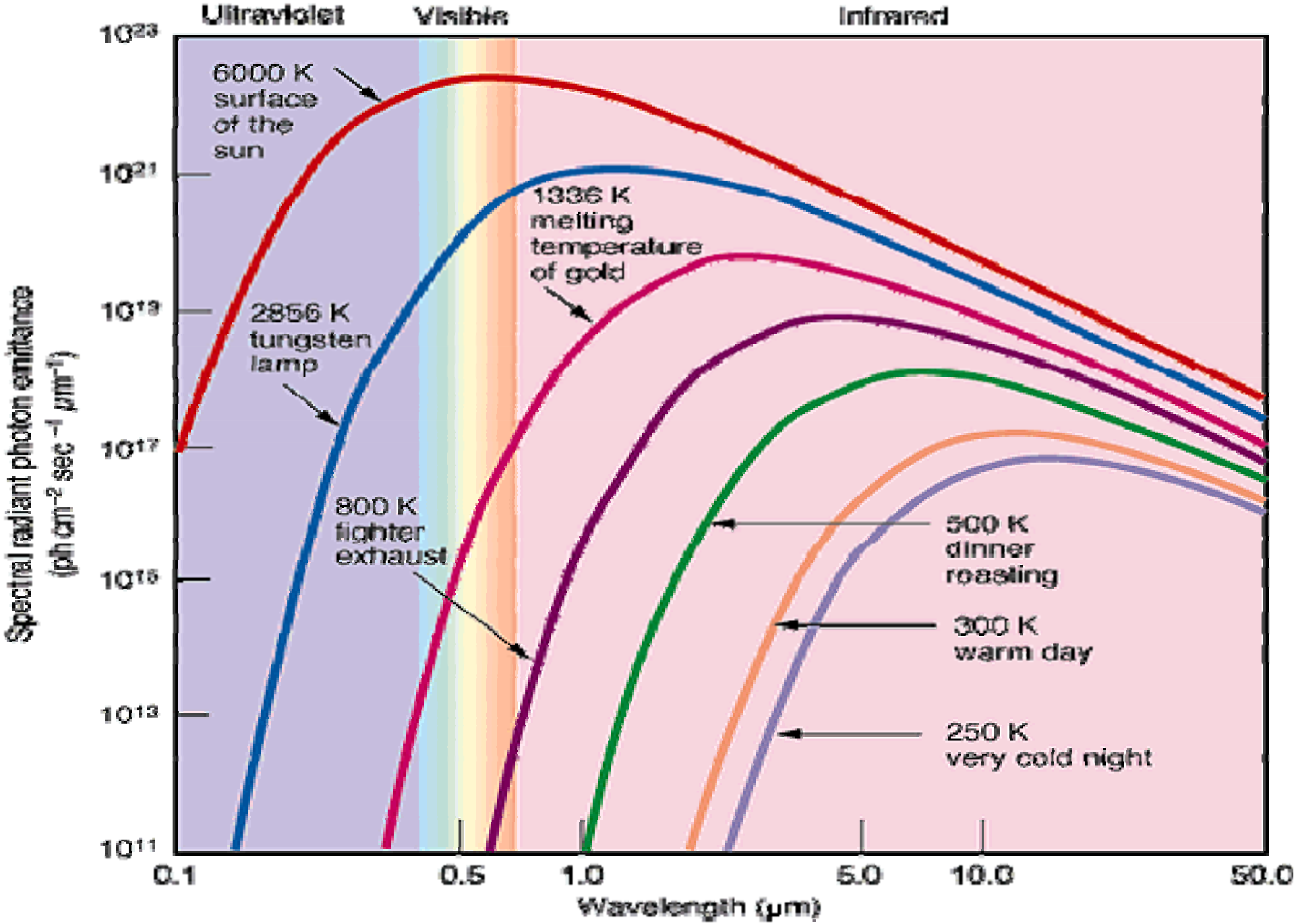
1. IEEE Press 2004 "Comp. Intel" Ch.16 Szu Unsupervised Learning ANN,
2. Shanghai Sci Ed Publ. 2003 Szu & Zhang" Intel Image Proc. Blind Sources Sep.

"From Tanks to Tumors" State of the Art: Healthy Breasts (shown left) and After 10 min. waiting (right) by one camera passive Thermal Scan



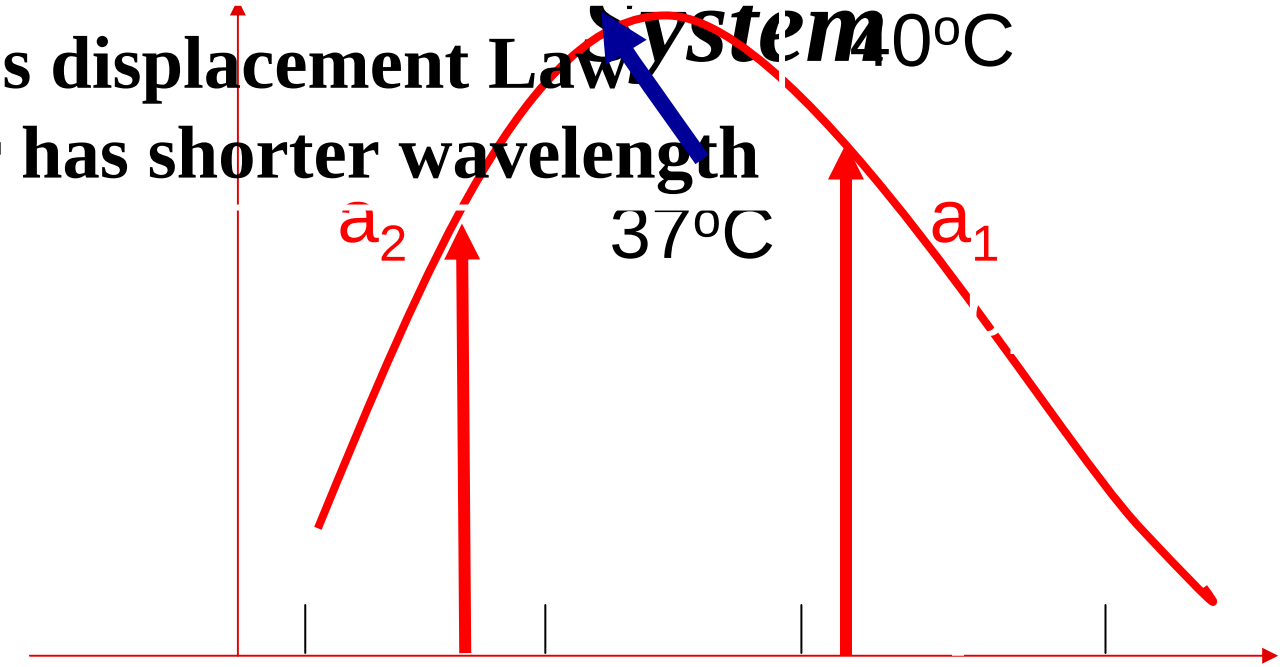
State of the Art: Pathological Breasts Before (shown left) and After 10 min. waiting (right) by one camera passive Thermal Scan (IRI)



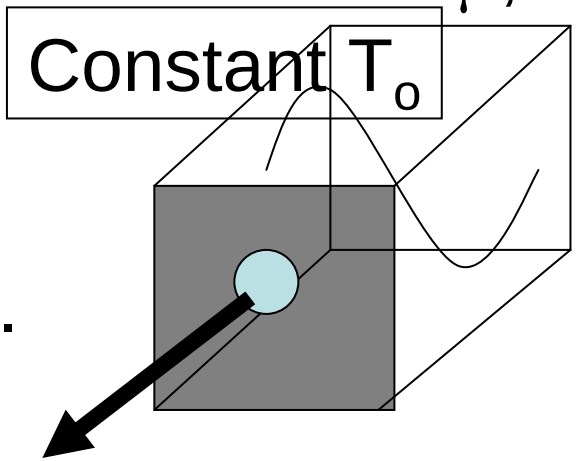

Ultraviolet
Visible
Infrared
6000 K surface of the sun
1336 K melting temperature of gold
2856 K tungsten lamp
800 K fighter exhaust
500 K dinner roasting
300 K warm day
250 K very cold night
Spectral radiant photon emittance (ph cm⁻² sec⁻¹ μm⁻¹)
10²³
10²¹
10¹⁹
10¹⁷
10¹⁵
10¹³
10¹¹
0.1
0.5
1.0
5.0
10.0
50.0
Wavelength (μm)

Isothermal Equilibrium Black Body Radiation Planck Quantum Mechanics System

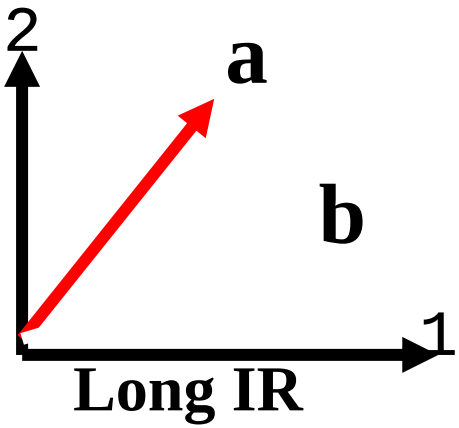
Wien's displacement Law: hotter has shorter wavelength

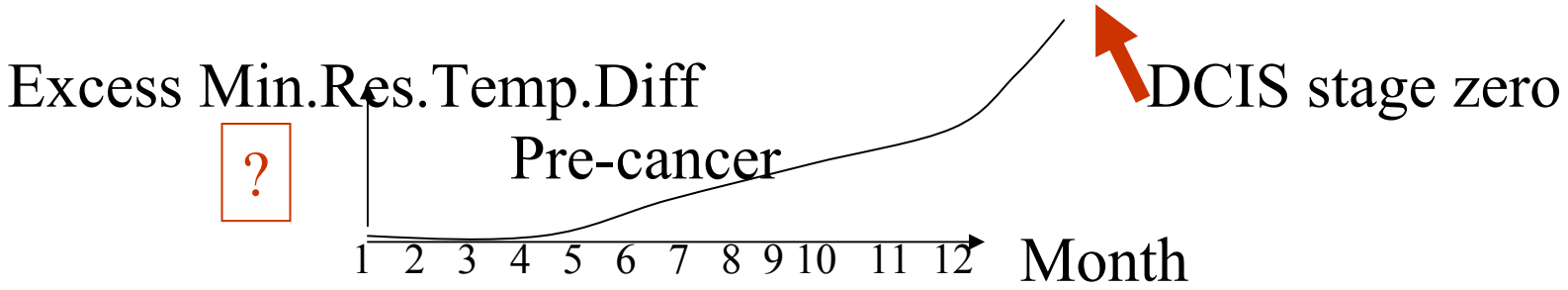


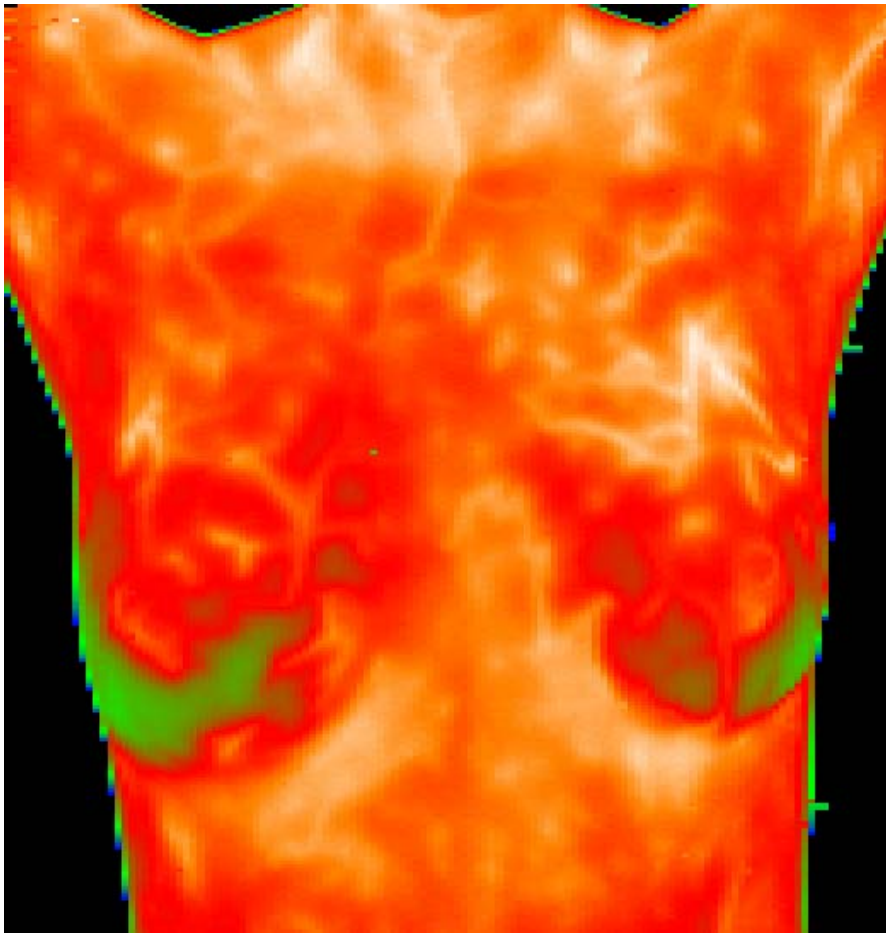
Visible Mid IR 3 to 5 μ;Long IR 8 to 12 μ

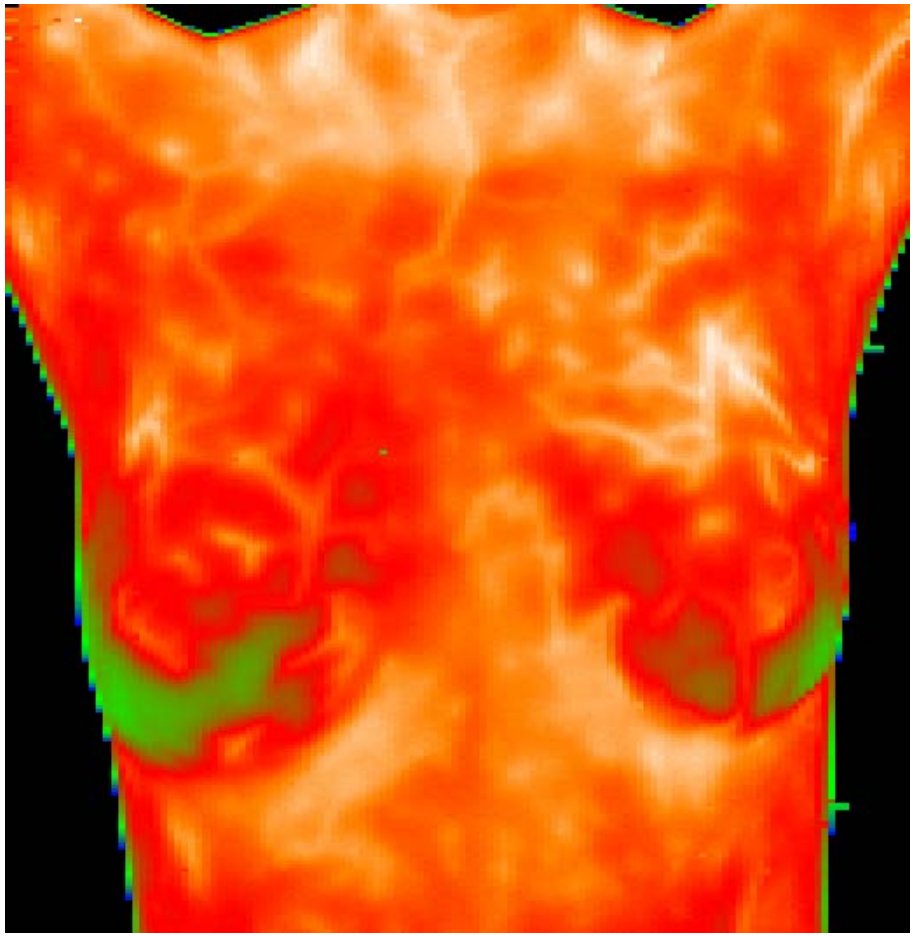


Blackbody Expt.







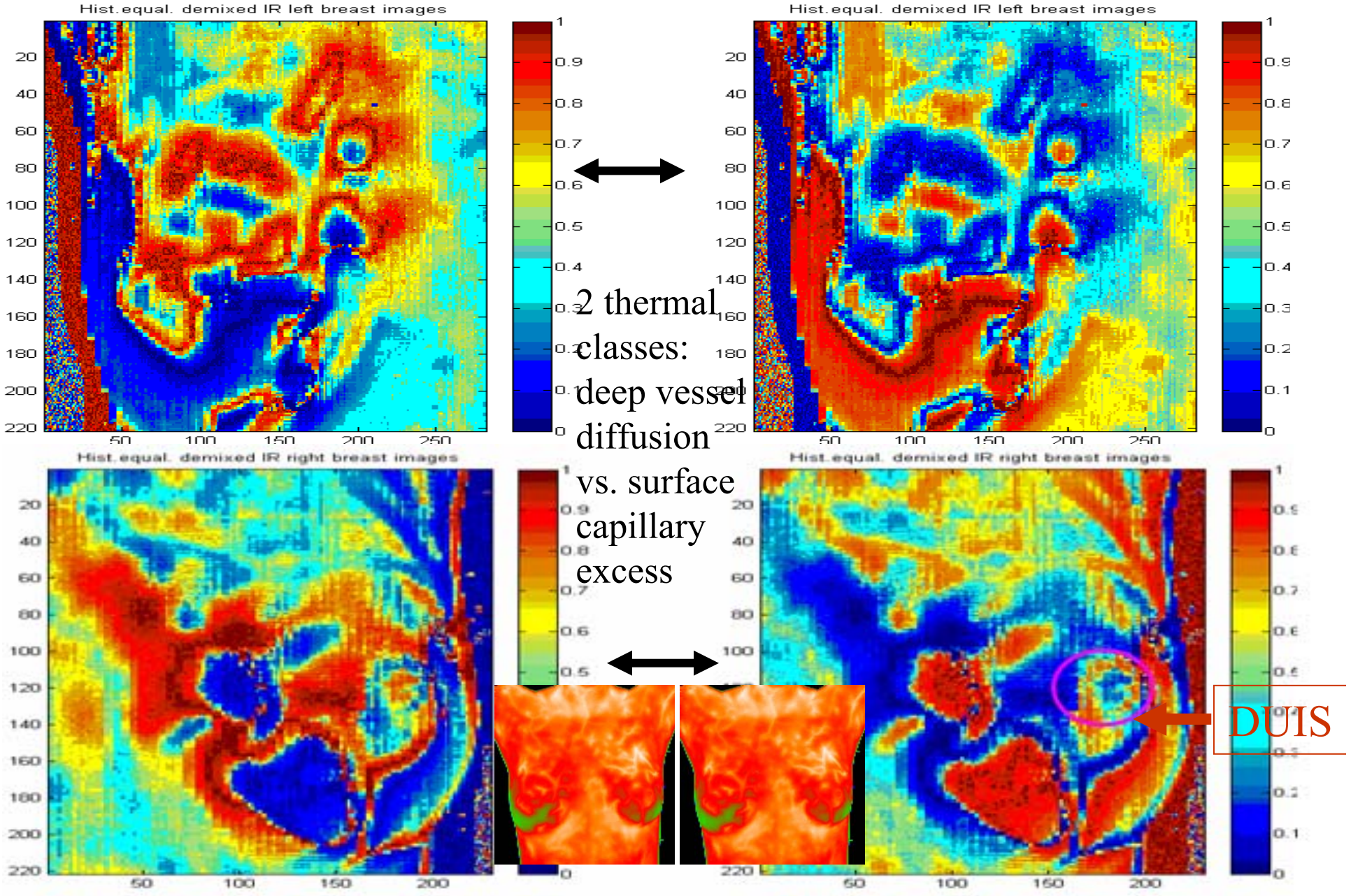


Mid IR

Long IR

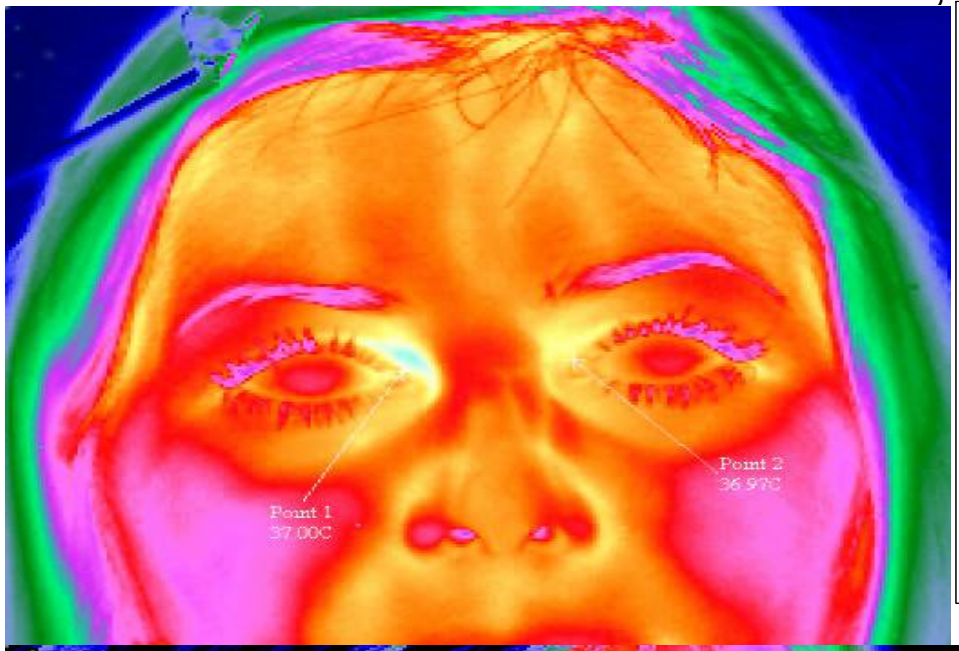
Two cameras passive imaging track risky patients without radiation hazard and waiting

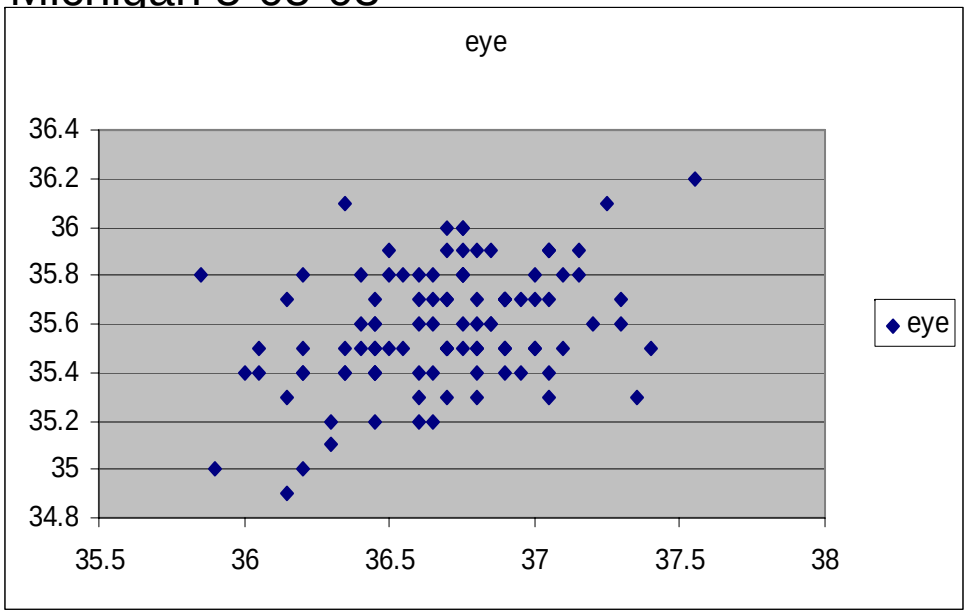
Reference: Christine Gorman," Rethinking Breast Cancer", ***Time Magazine***, p.p.50-58, Feb 8, 2002

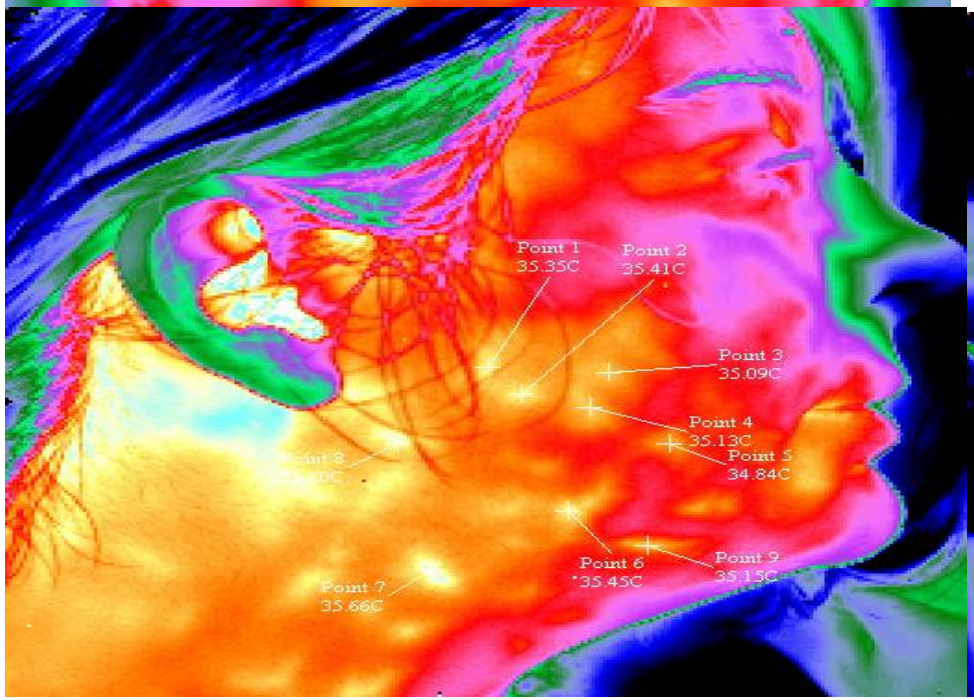

Hist.equal. demixed IR left breast images
Hist.equal. demixed IR left breast images
Hist.equal. demixed IR right breast images
Hist.equal. demixed IR right breast images
2 thermal classes: deep vessel diffusion vs. surface capillary excess
DUIS

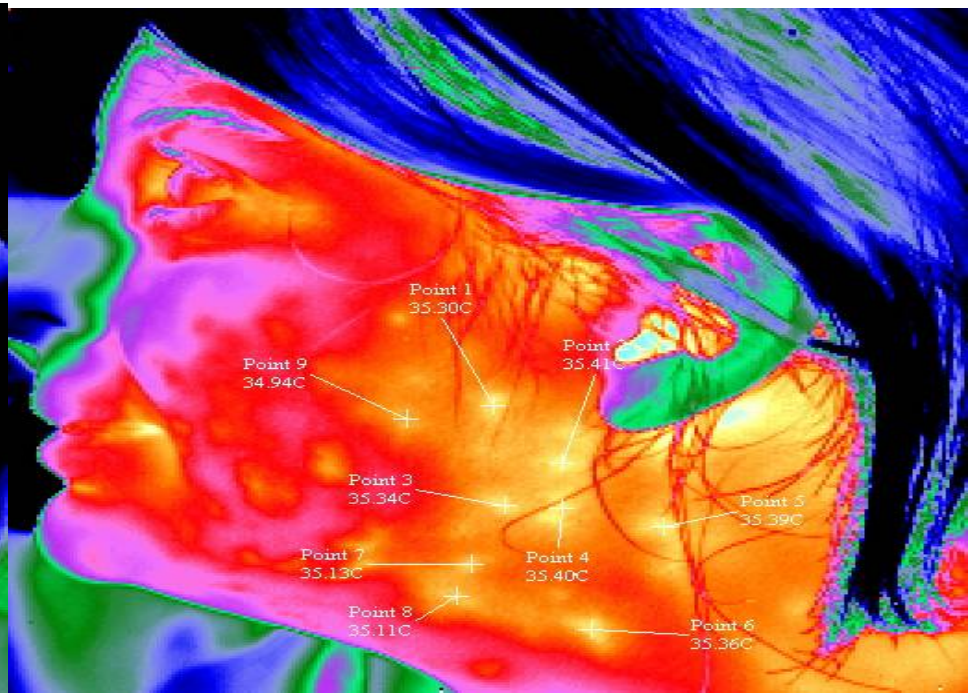
EXPERTS RE-DEFINE SARS AIRPORT THERMOGRAPHY SCREENING

Detroit/Toronto , Michigan 5-05-03









3. Nano Camera Fabrication Technology

Preventive diagnoses must be done regularly at homes.

We need to design non-cryogenic-cooled consumer electronics camera called “(passive non-contact accurate to 0.02 degree) dual infrared) Spectrum Thermometer ™” replacing traditional contact thermometers based on either thermal coupler or Mercury thermometer

Biomimetic Fovea Camera by means of Carbon Nano Tubes overlayed CCD array

(1) Implementation issue:

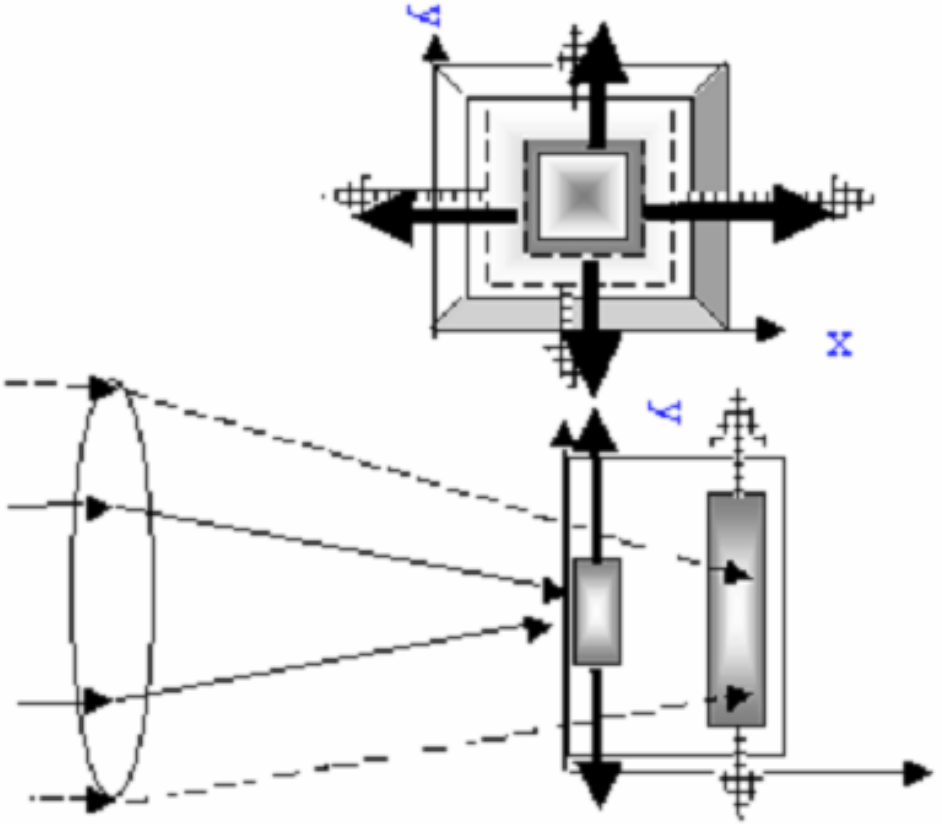
nano-robot

(2) Algorithm issue:

unsupervised Fusion

Non-Cryogenic Co-Axial Fovea Design of Infrared Two Color Pixels Planes*

- full-band IR lens focused at dual focal planes along uniaxial for both MidIR &LongIR
- Band-selective Carbon NanoTube (CNT) generates only 1% or less occlusion over Long IR PFA CCD pixel.
- CNT might only suffer 1D noise that permits a non-cryogenic cooling.
- Un-cool FPA CCD for LongIR & steering & pointing at ROI
- Unsupervised fusion for unbiased feature extraction.



Uncooled night vision

CCD @$1K

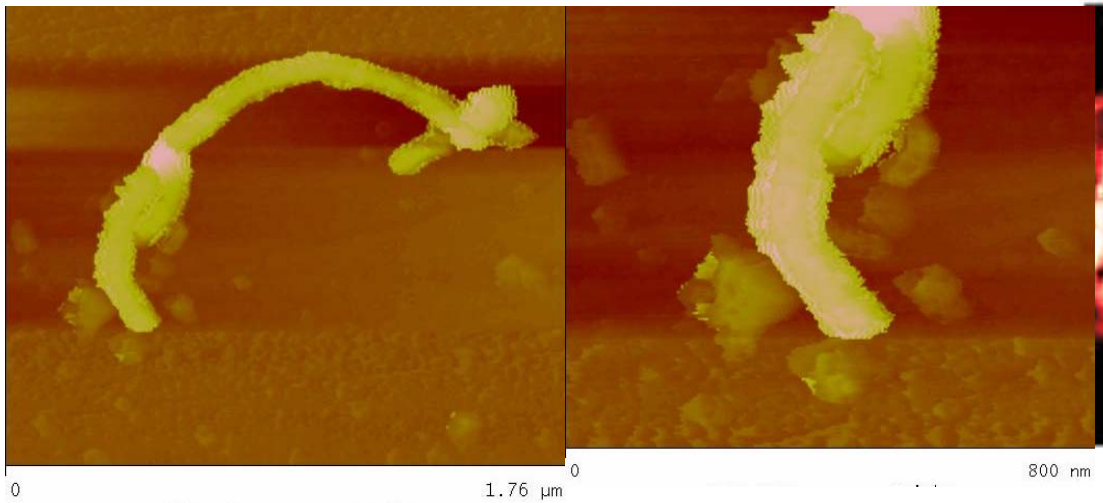
*Szu et al. Patent Disclosure 2004

Non-cryogenic Spectrum Thermometer based Mid-IR 1-D CNT & Long-IR FPA

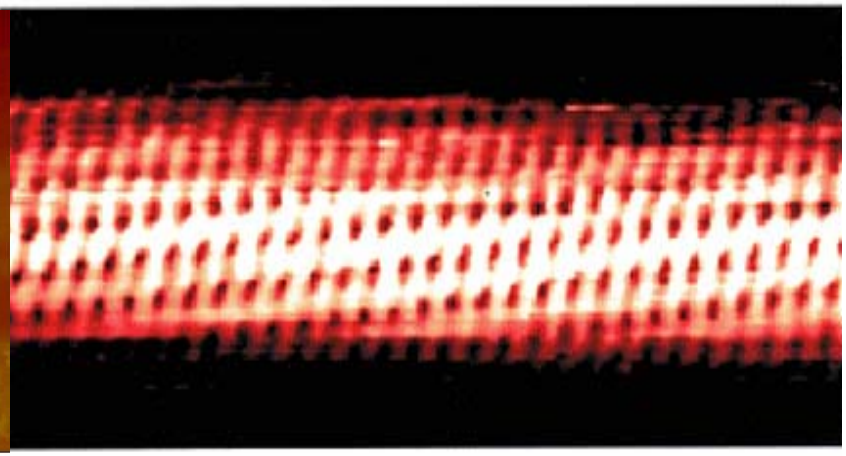
- Navy Patent 2004 unsupervised fusion by Szu et.al. can provide unbiased feature extraction.
- Un-cooled 2 Color Infrared Camera under $1K can provide mini-UAV new capability and dual usage for family spectrum thermometer for sport medicine and early tumor diagnoses.

CNT (S. Iijima, CNT, Nature 354,56,1991; Nobel Prize: C_{60} Bucky, Rice U. Rich Smoley 1989)

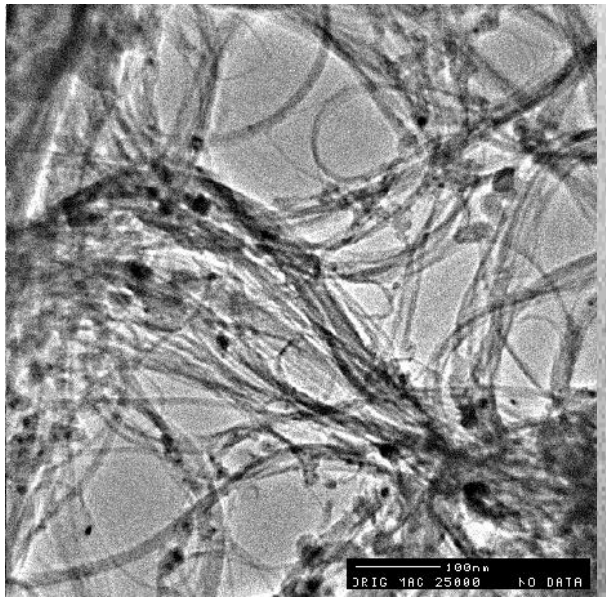
AFM image



STM image of CNT

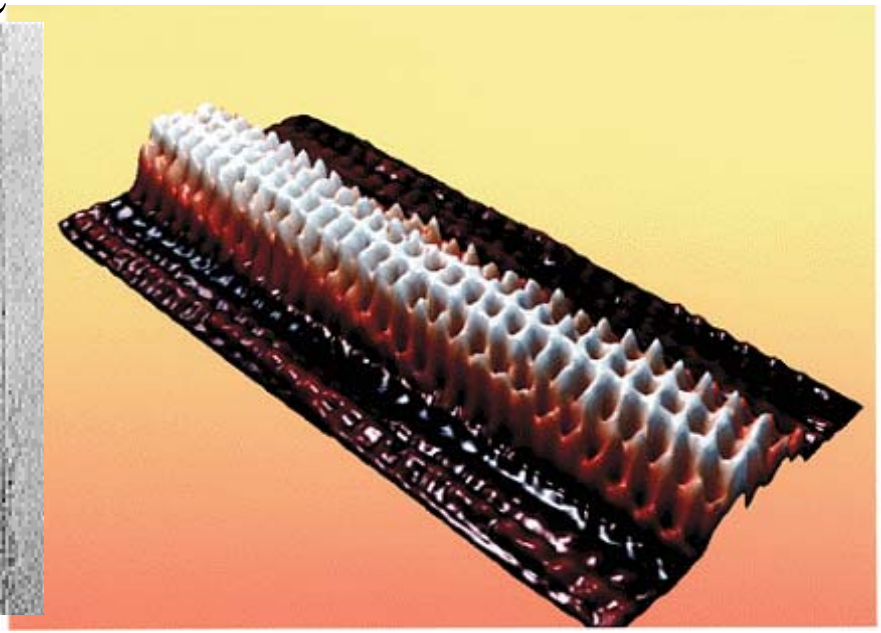


SEM image



TEM image

Cees Dekker, Delft Univ of Tech

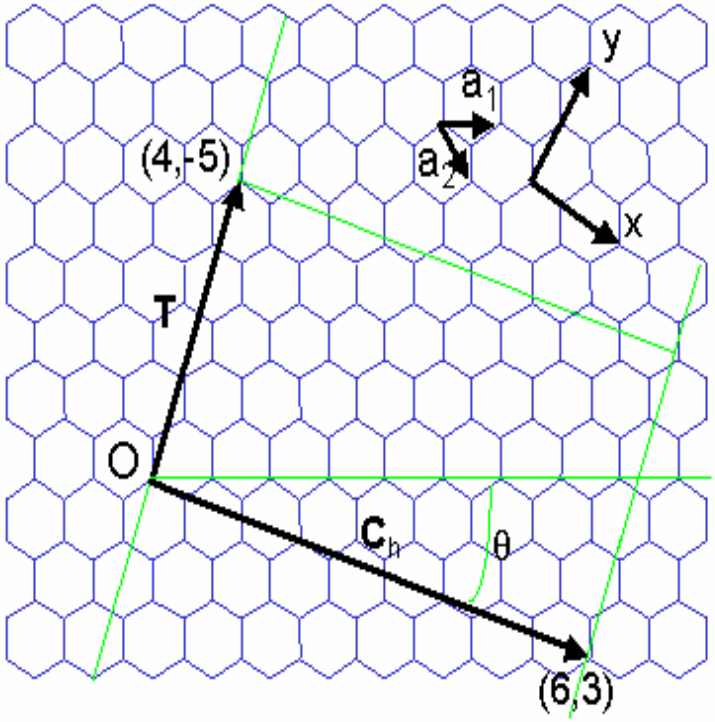


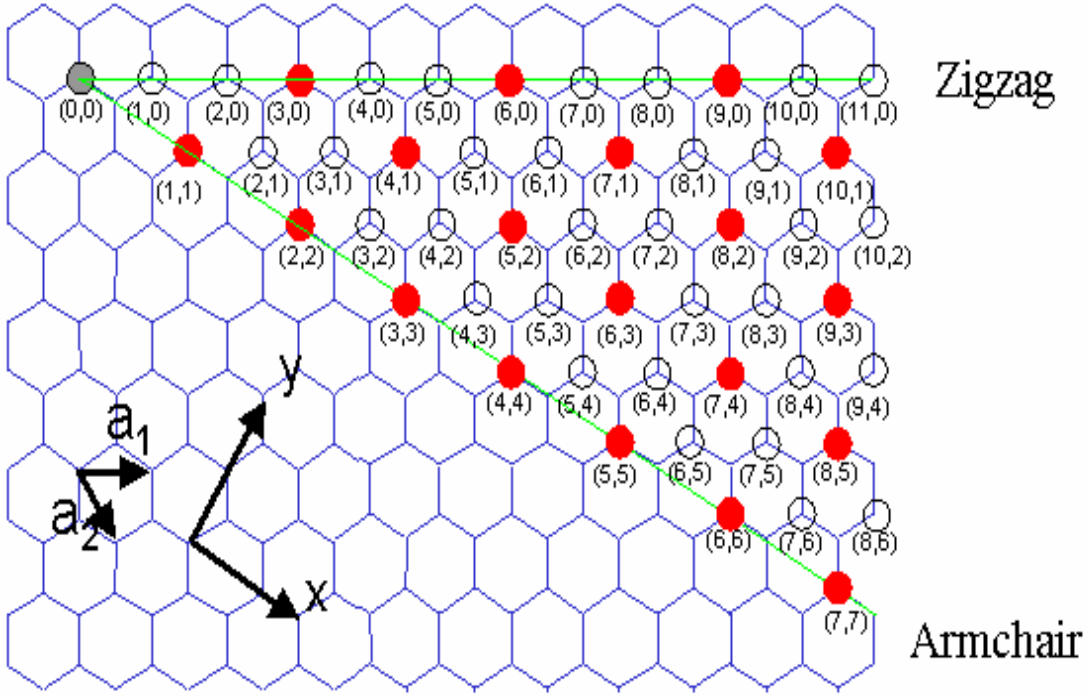
Why Nano technology CNT for double color Infrared Camera?

- Nano 10^{-12} meter is a small quantum system of molecule about 10 Angstrom (10 Dalton)
- It can fit in small space for local action.
- It is at the border of classical physics and quantum mechanics
- It can be either conductor or semiconductor depending density of states exist at Fermi surface or not.
- As a conductor, it has a sharpest field gradient for point discharge for the ionization of neutral gas
- It enjoys limited quantum phonon noise

CNT Index Scheme

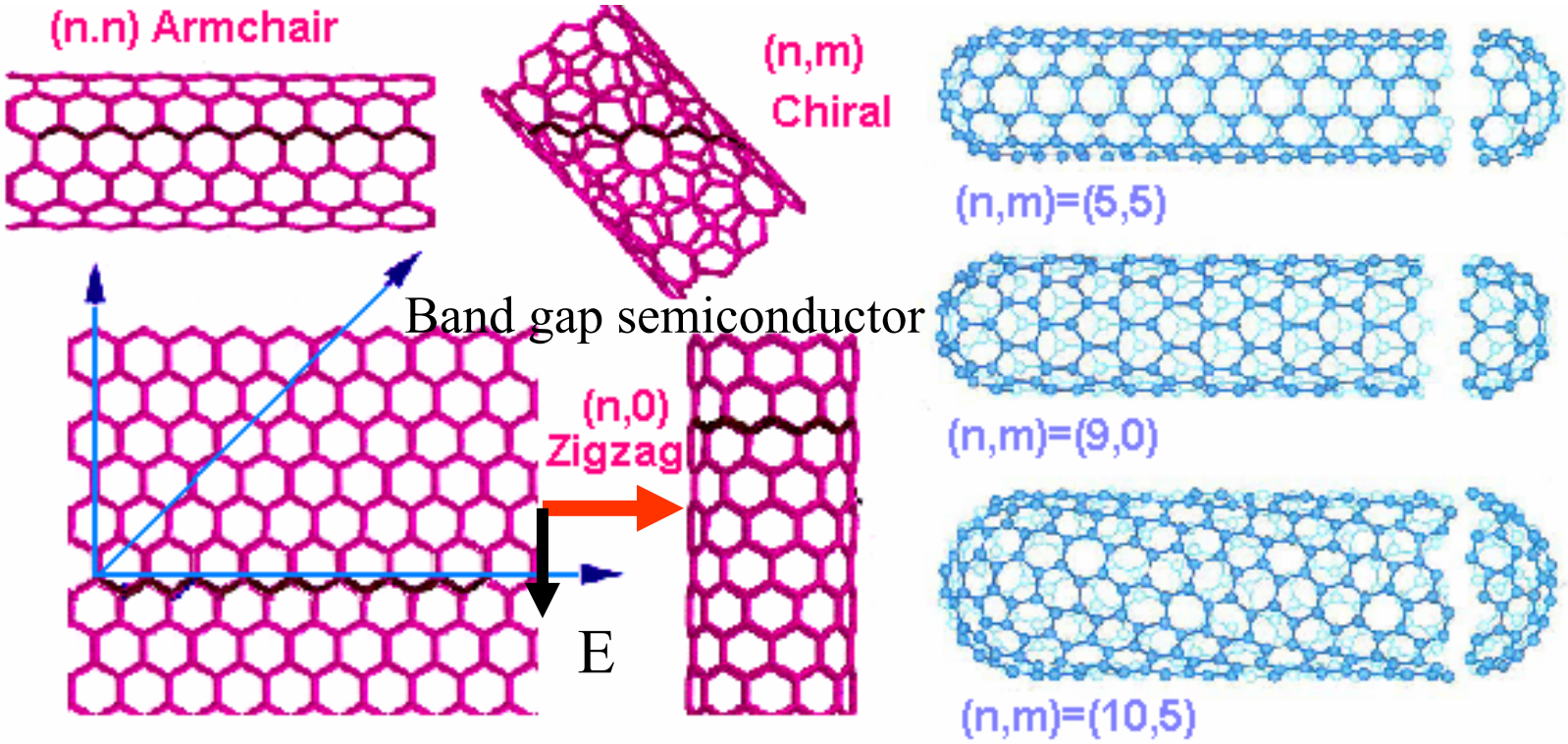
- Folding Chiral vector is defined as $C_h = n\vec{a}_1 + m\vec{a}_2$
- The translation vector T is perpendicular to the chiral vector
- (n,m) is the index of CNT





Classification of CNT

- Two major categories : Conductor Armchair (n=m); semi-conductors Zigzag (n=0 or m=0 non-multiple of 3).
- Single-Wall; Multiwall



Dispersion Relation

Saito et.al. from MIT (1992 *APL*, *p*.2204)

$$E_{2D} = \pm\gamma[1 + 4\cos(\frac{\sqrt{3}}{2}k_x a)\cos(\frac{1}{2}k_y a) + 4\cos^2(\frac{1}{2}k_y a)]^{1/2}$$

$$\vec{C}\vec{k} = 2\pi \,\text{int}\, egers$$

$$Conductors\, E_F = 0 \;\; where \;\; density \;\; of \;\; states \neq 0$$

$$n_1 = n_2$$

$$2n_1 + n_2 = 3x \,\text{int}\, egers$$

Dispersion Relation: Wave Eq. exp(iωt-kx)

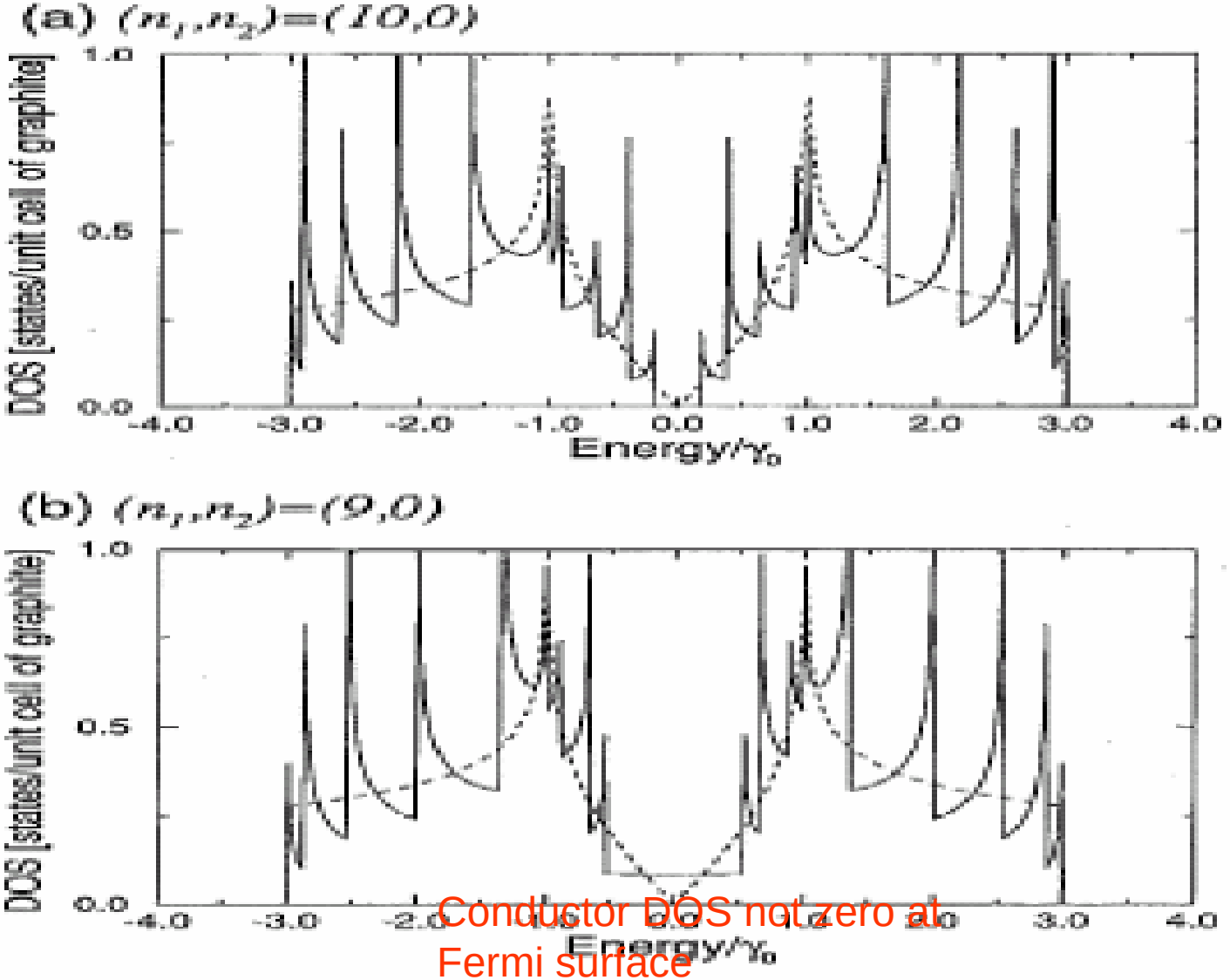


FIG. 3. Electronic density of states for two (n_1,n_2) zigzag fibers: (a) (10,0) and (b) (9,0).

Single photon SNR Quantum eff.

CNT bandgap at Mid IR 3 to 5 micrometer

$$Signal\ photon\ \Delta E = \hbar\omega = h\frac{c}{\lambda} = 0.414eV \Leftrightarrow 0.248eV$$

$$Between\ room\ temperature\ T = 300^{o}K;\ Liquid\ Nitrogen\ T = 77^{o}K$$

$$Gaussian\ noise\ energy\ K_B T = \tfrac{1}{40}eV = 0.025eV;\ 0.006eV\ a\ factor\ 4$$

$$Johnson\ \ shot\ noise\ whose\ mean = \text{var}iance$$

$$1D:\ \ \frac{1}{2}K_B T < dark\ current < 3D:\ \frac{3}{2}K_B T\ at\ room\ temperature$$

Structure Dependency of Bandgap

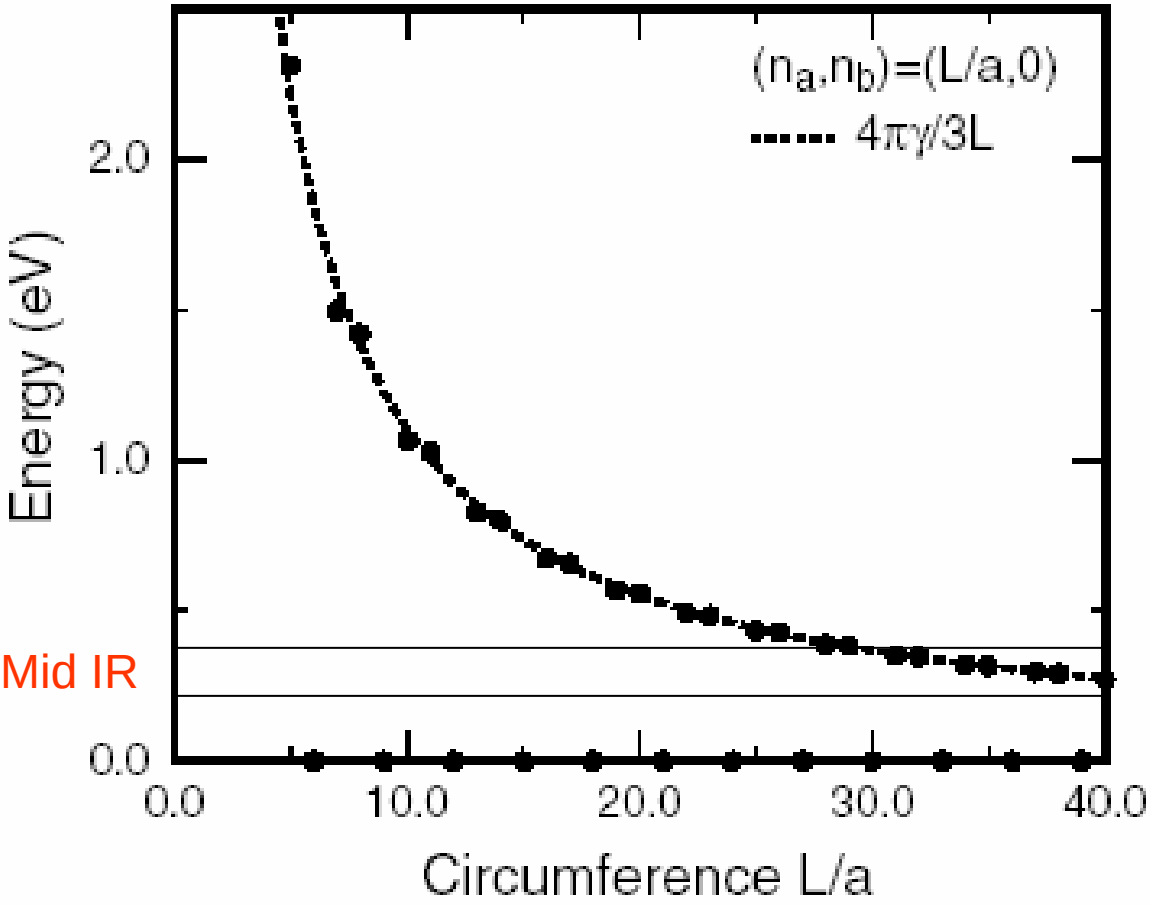
Mid IR 3 to 5 micrometer

$$Signal\ \ photon\ \Delta E = \hbar\omega = h\frac{c}{\lambda} = 0.414eV \Leftrightarrow 0.248eV$$

(Bandgap of zigzag CNT)

- Armchair (n,n) nanotubes are metals

- $(n, n+3i)$ tubes (with an integer) are small gap semiconductors with $E_{gap} \propto 1/R^2$
- Other tubes have larger gaps proportional to 1/R



Computation of Bandgap

- Single-wall (CW) CNT

$$E_g = \frac{2\gamma_0 a}{D}$$

γ_o =2.6eV: pp π hopping interaction,
a=1.41Å is the C-C nearest neighborhood distance,
D is the diameter of the SWCNT

- Multi-wall CNT

$$E_g \approx \frac{3}{\sqrt{2}} M_i \varpi^2 \frac{a^3}{D}$$

$M_i \approx 2x10^{-23}$ is the mass of carbon atom,
$\varpi \approx 1600cm^{\wedge}-1$ is the characteristic phonon frequency

Grand Challenge:
Un-cooled Mid IR SNR 5 orders of Magnitude

Time integration of 100 signal photons is needed

$$1-D\ dark\ current\ \frac{1}{2}KT << 3-D\ dark\ current\ \frac{3}{2}KT$$

$$0.0125eV < dark\ \ current < 0.0375eV$$

$$SNR_{room} = \frac{0.4 \Leftrightarrow 0.2}{0.01} = 40 \Leftrightarrow 20\ if\ 1-D\ (otherwise: 13 \Leftrightarrow 7\ for\ 3-D)$$

$$at\ 77^{o}K \Rightarrow noise\ 0.006x3/2 \cong 0.01\ \ (1\%eV)$$

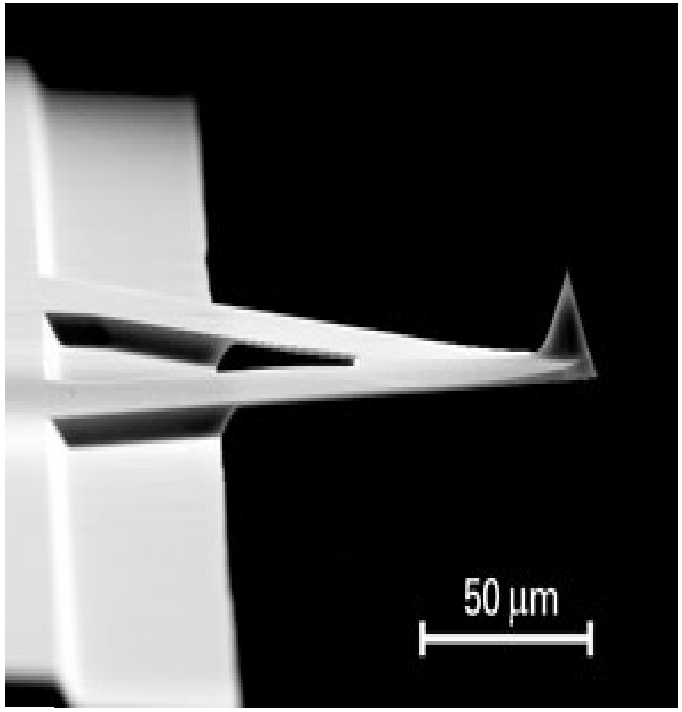
$$SNR_{cryogenic} = \frac{0.4 \Leftrightarrow 0.2}{0.01} = 40 \Leftrightarrow 20\ if\ 3-D$$

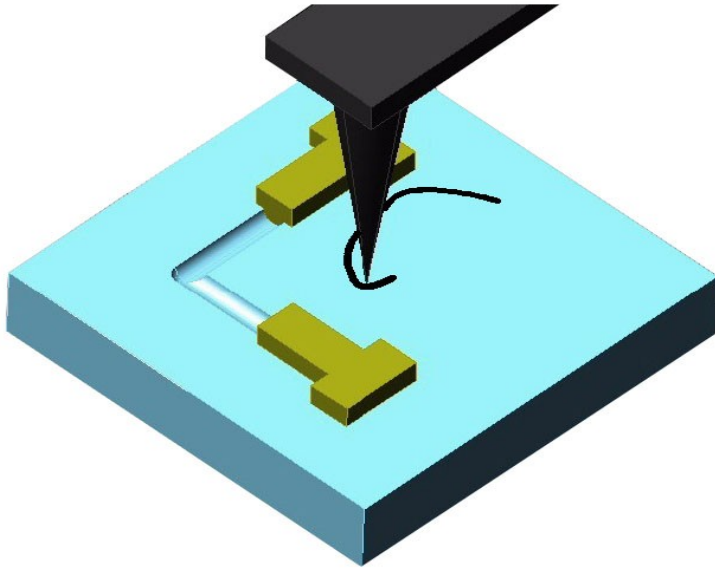
Classical Nanostructure 1/f noise vs. Quantum CNT Lorentz noise

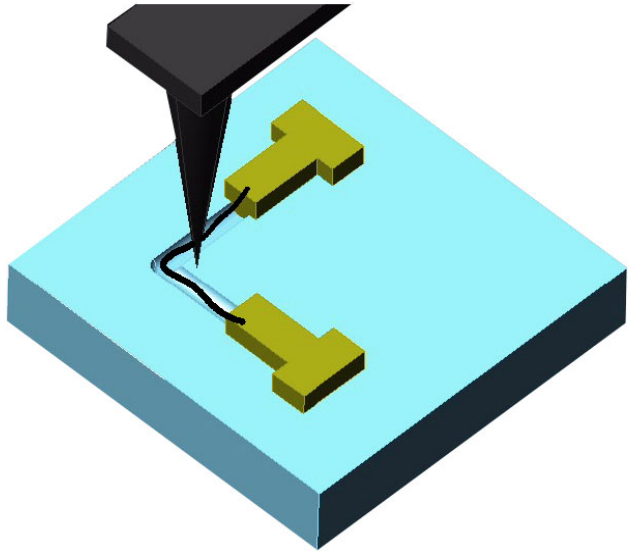
- Wiener-Khtchine Theorem $FT\{psd(f)\}=<c(x_o+x)c(x_o)>$
- Nanowire, e.g.GaAs; FT{1/f} = Step(x)=1 x>0 all correlation scale i.e. self-similar noise characteristics
- CNT: $FT\{[1+(f/a)^2]^{-1}=\exp(-|x|/a)$ where $a=1.41A^o$

Fabrication of Sensor Pixels

- AFM based nanoassembly
- Pushing Nanotube into nanofixture
- Tuning CNT band gap by modifying its shape

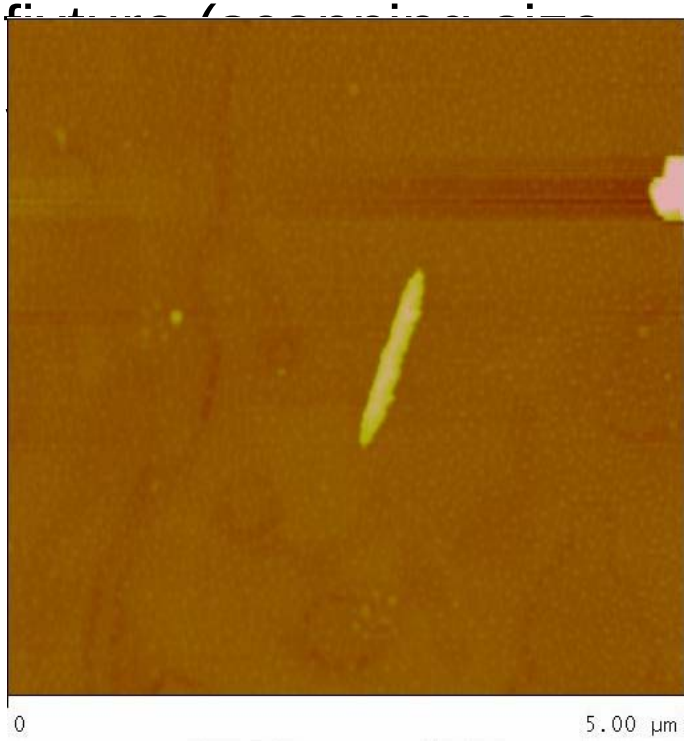




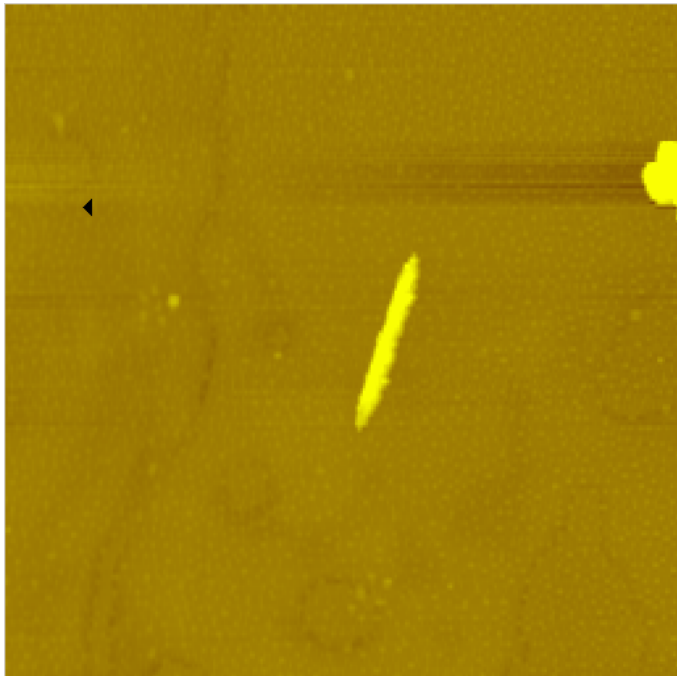


Nanoassembly Experiment

- Creating a fixture and then pushing a 100nm silver nanowire into the



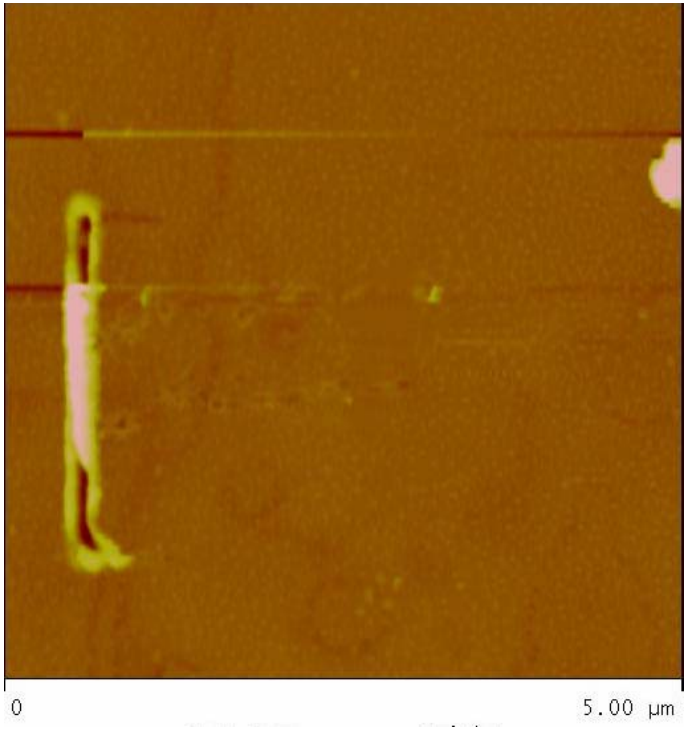
AFM image



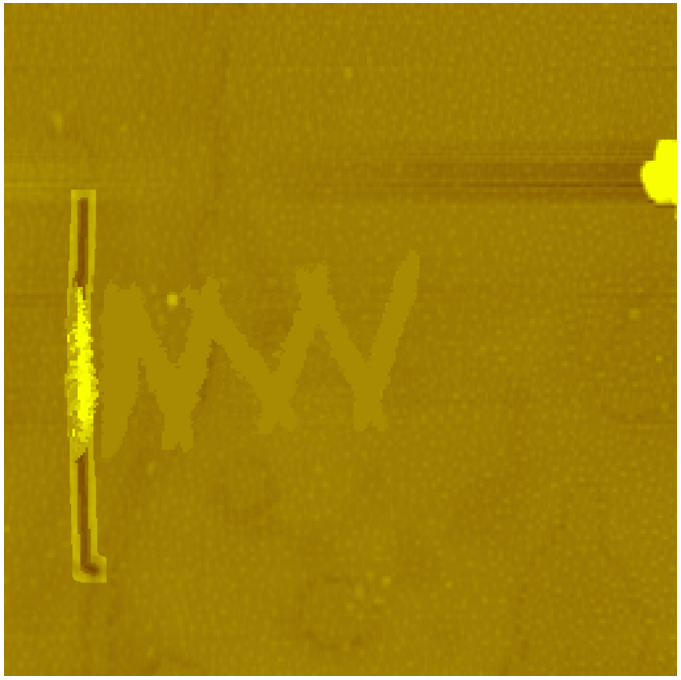
Real-time visual display

Nanoassembly Experiment

- Creating a fixture and then pushing a 100nm silver nanowire into the fixture (scanning size 5um)



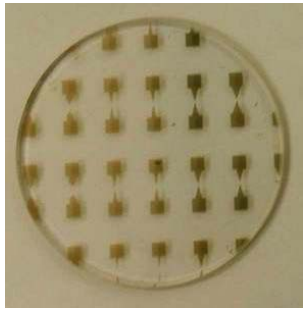
AFM image

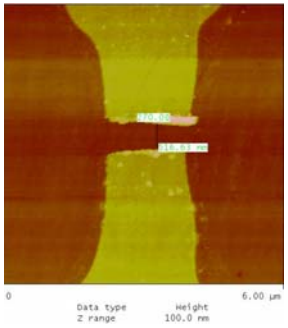


Real-time visual display

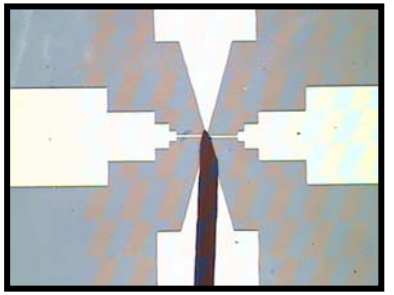
Fabrication Process

Substrate Fabrication



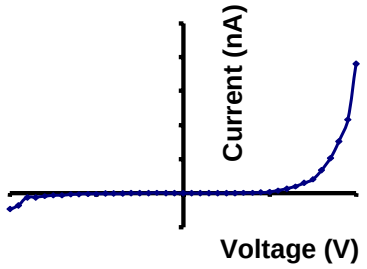


Nano lithography



CNT Deposition

CNT Assembly



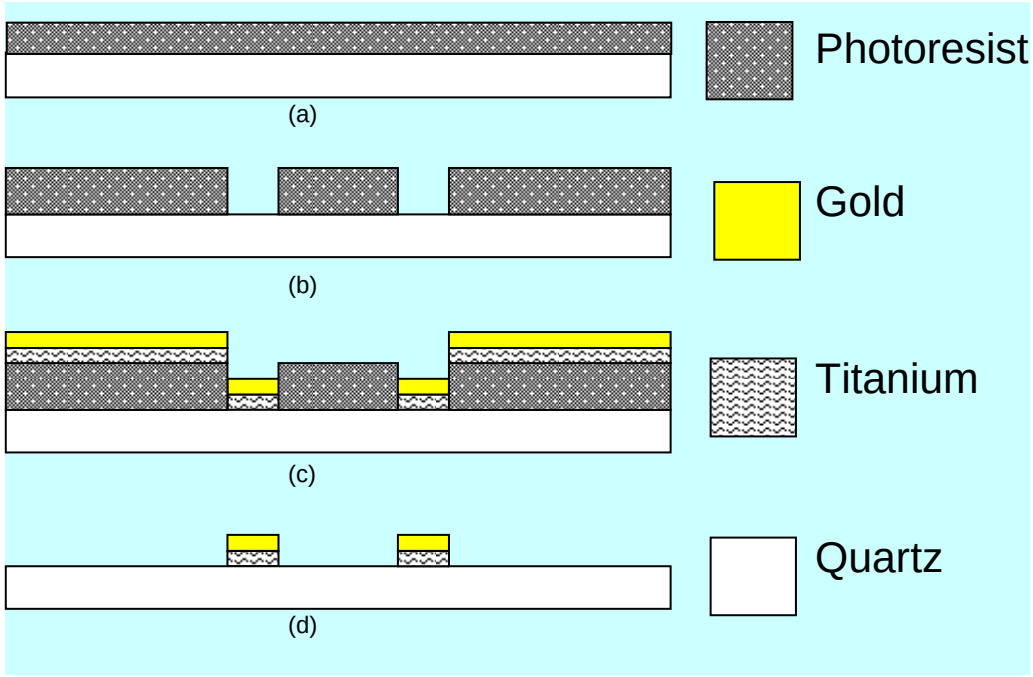
Band Gap Tuning

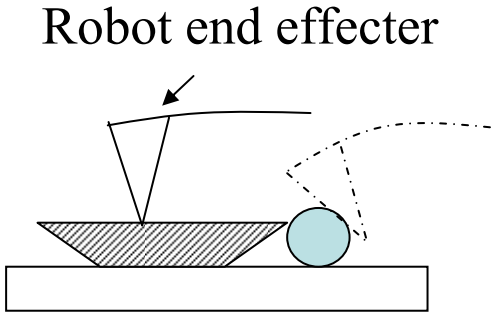
Chip Packaging

Final Testing

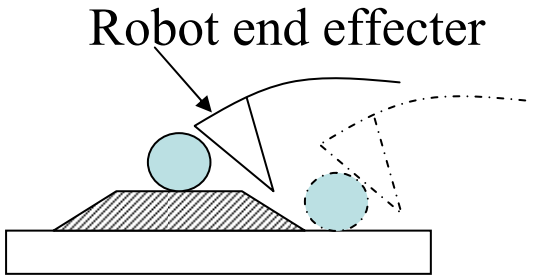
Substrate Fabrication

The electrode patterns are made from a lift-off process





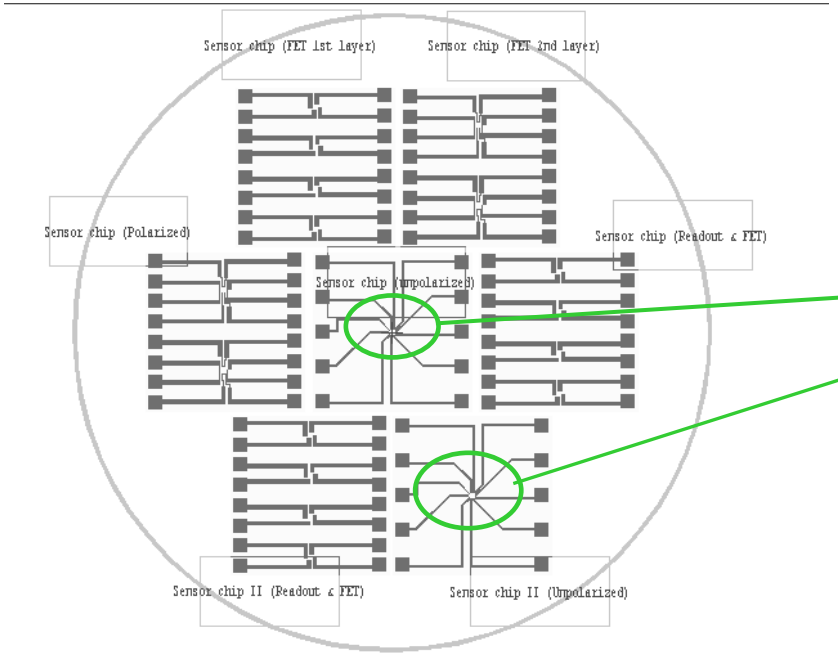
(a) etching process



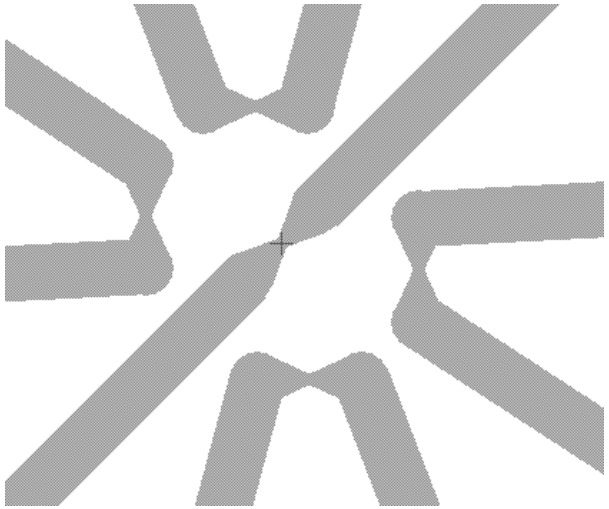
(b) lift-off

Mask Design and Fabrication

Unpolarized single pixel, which consists of five sub-pixel, is designed. The signal of the unpolarized single pixel is irrelated to the incident direction of the infrared.

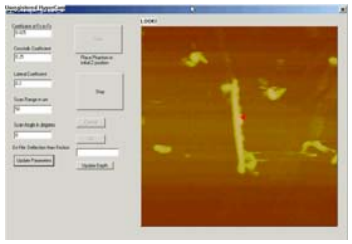


unpolarized single pixel



Fabrication and Assembly by Nano Robot

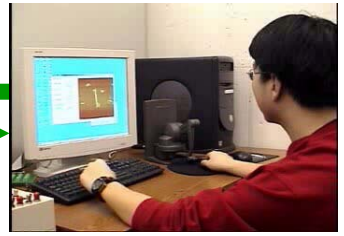
AFM Based Robotic System

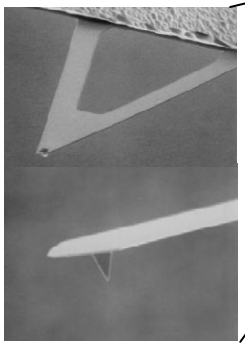


Real-time Videolized AFM Image Display

Control Command

Force and Image Info



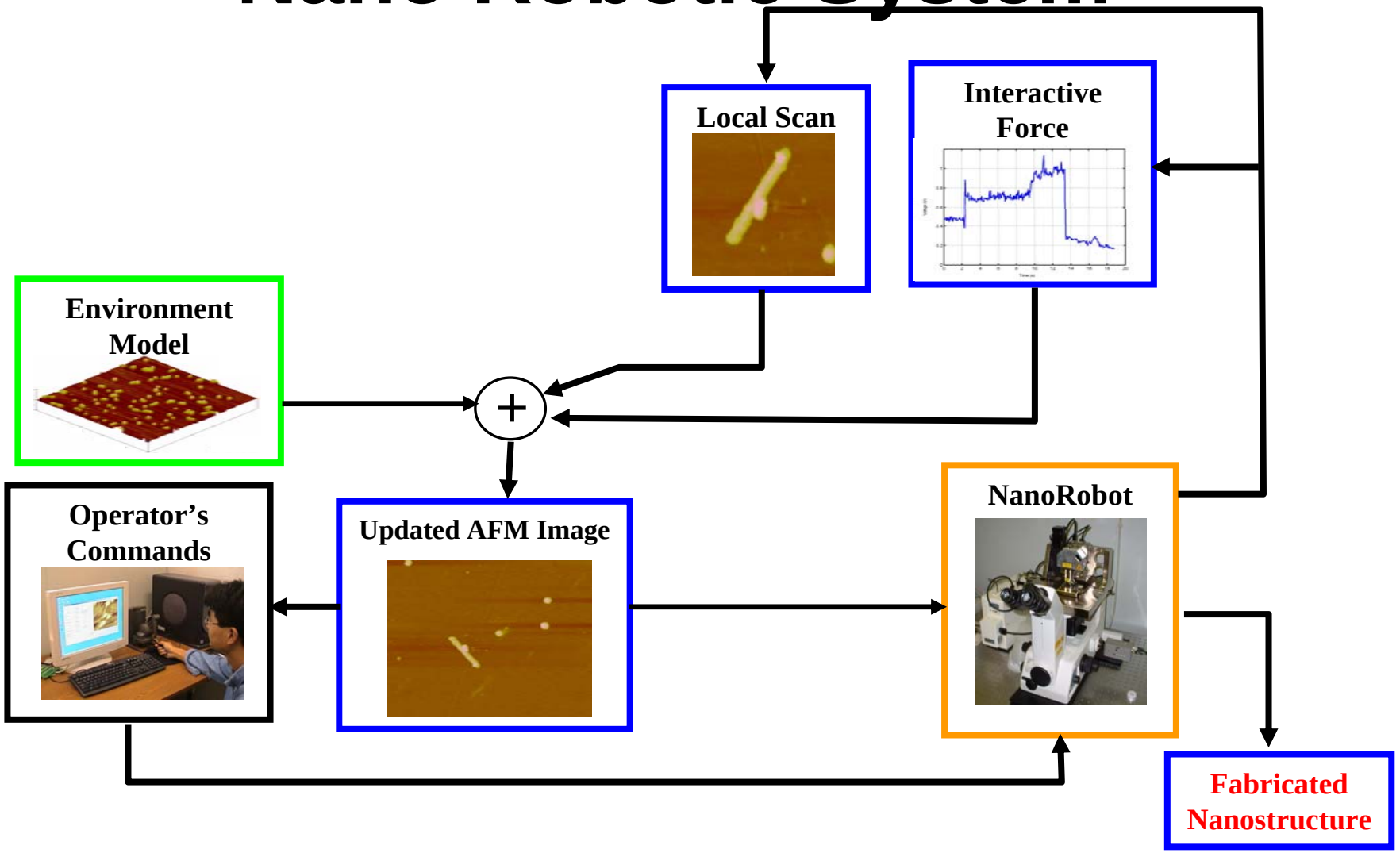


Robot End-Effector

Haptic Feedback

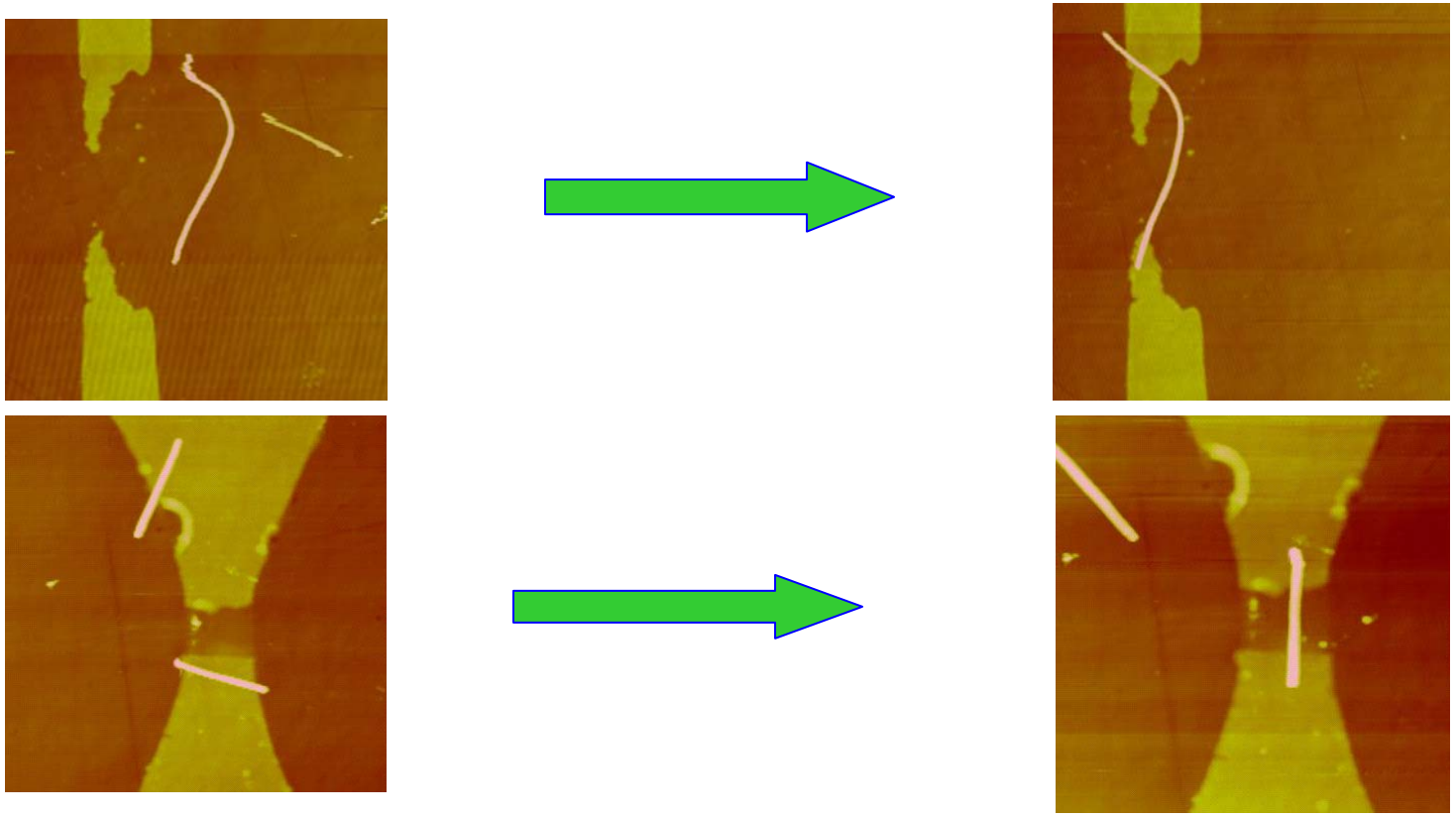
Command Generator

Augmented Reality Enhanced Nano Robotic System



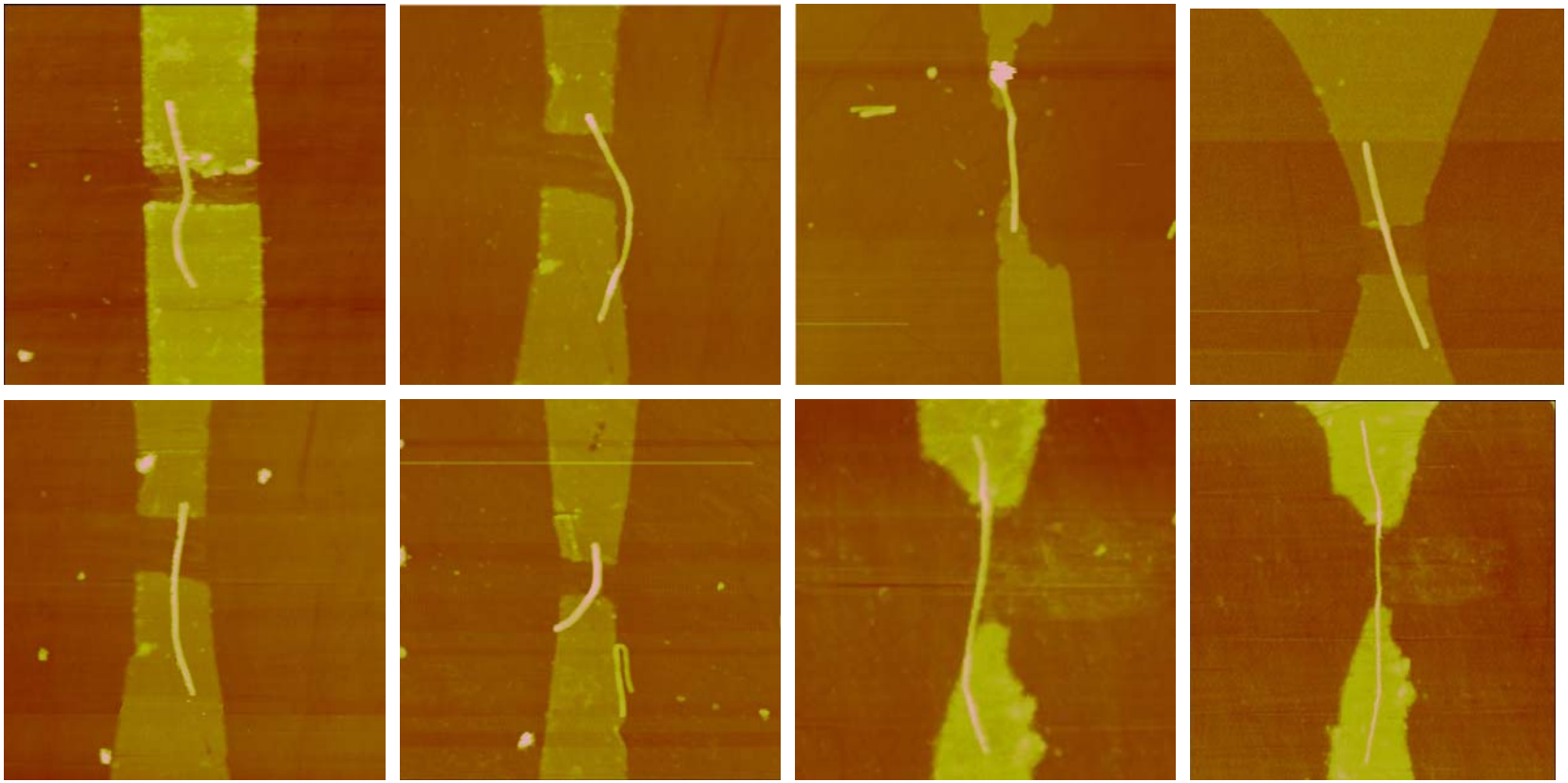
Placement of CNTs between Electrodes

The placement of a CNT between electrodes can be easily achieved using the AFM based nano robot



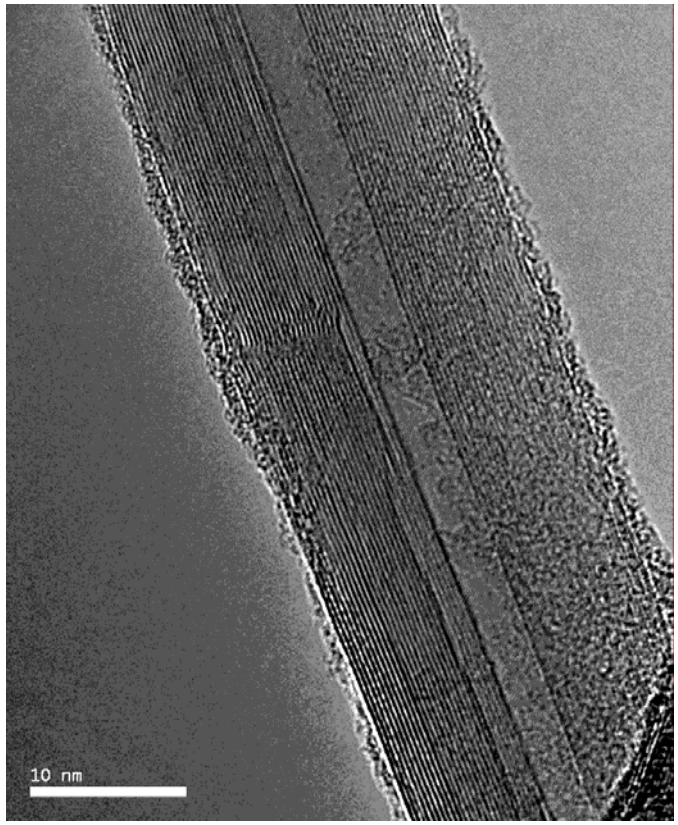
CNTs between Electrodes

- Assembled CNT based detectors

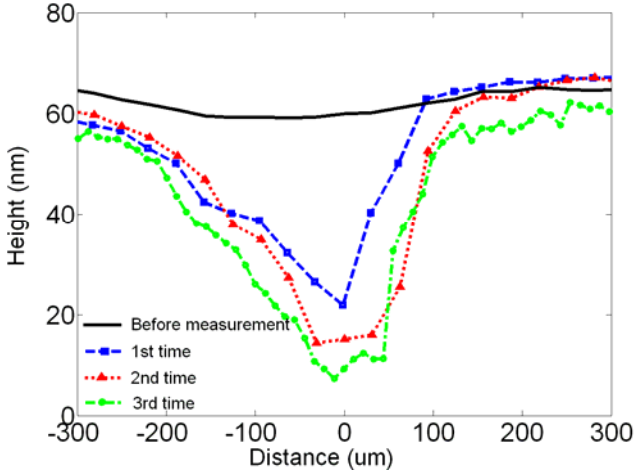


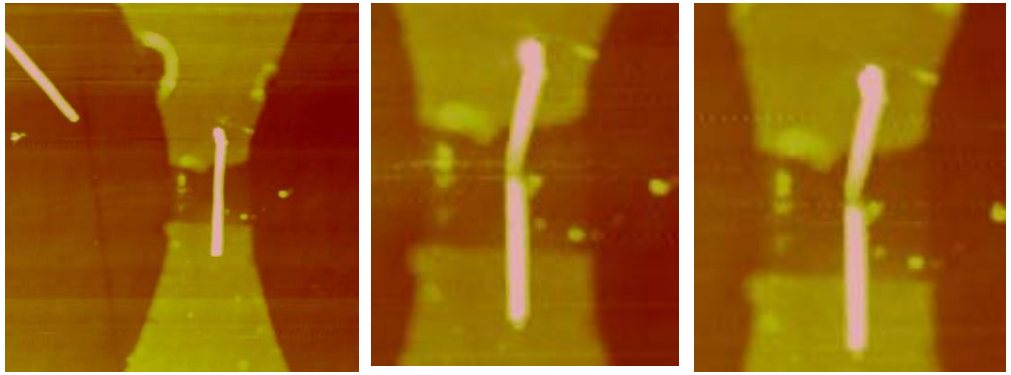
CNT Band Gap Tuning by Breakdown

TEM image of MWCNT

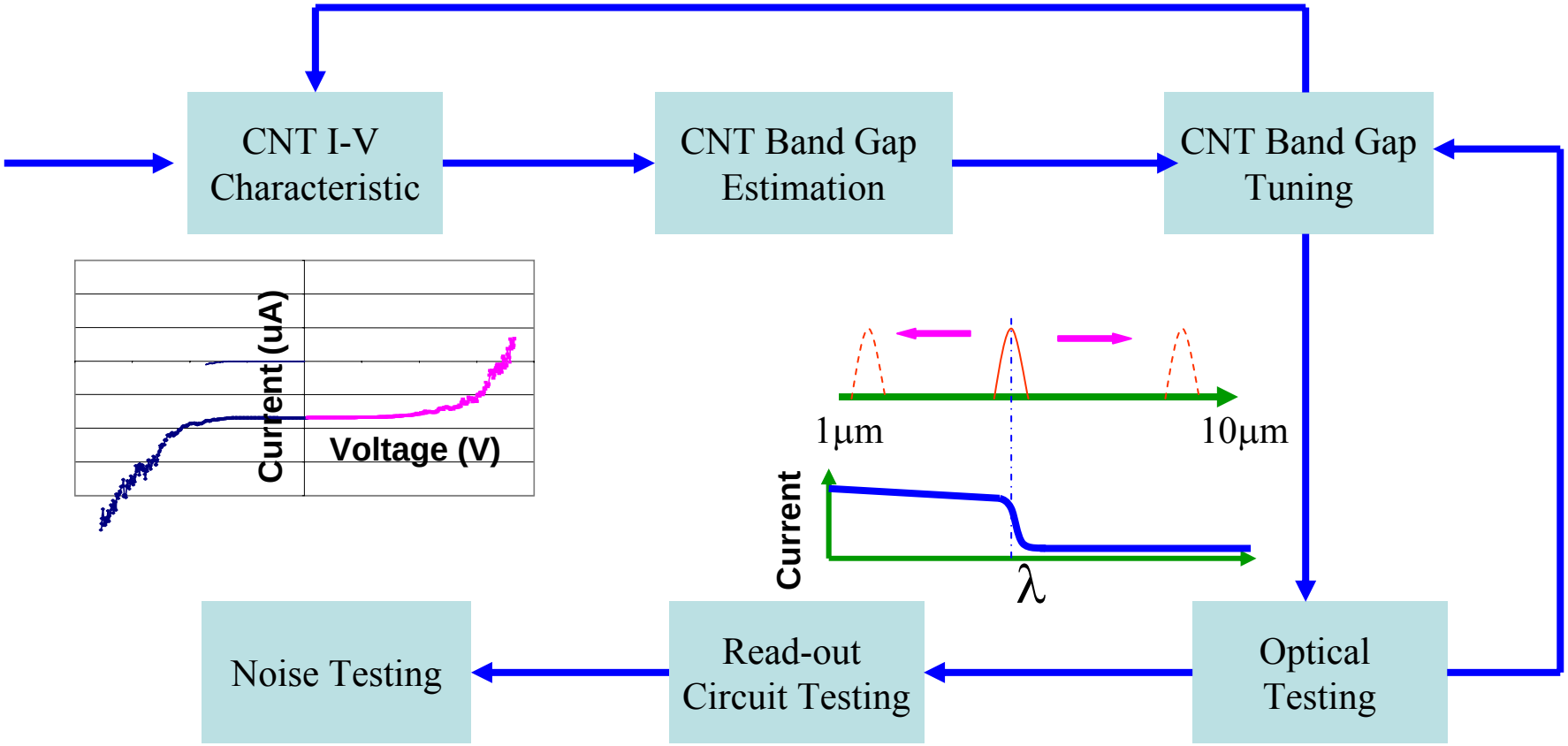


Topography of MWCNT



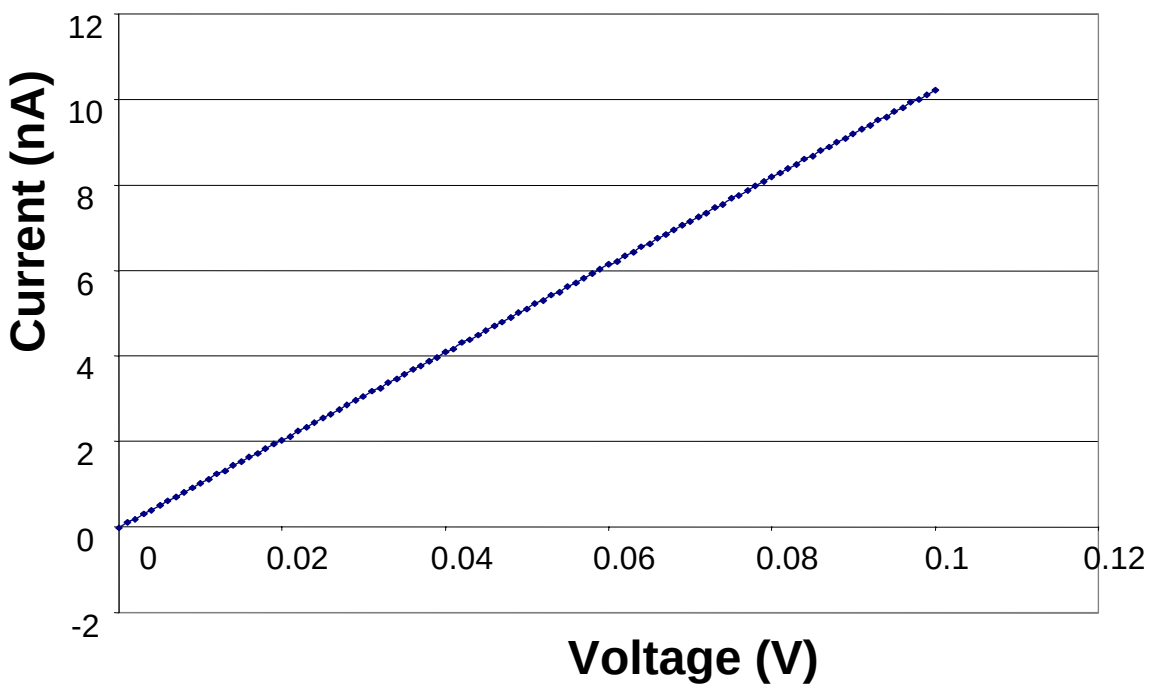


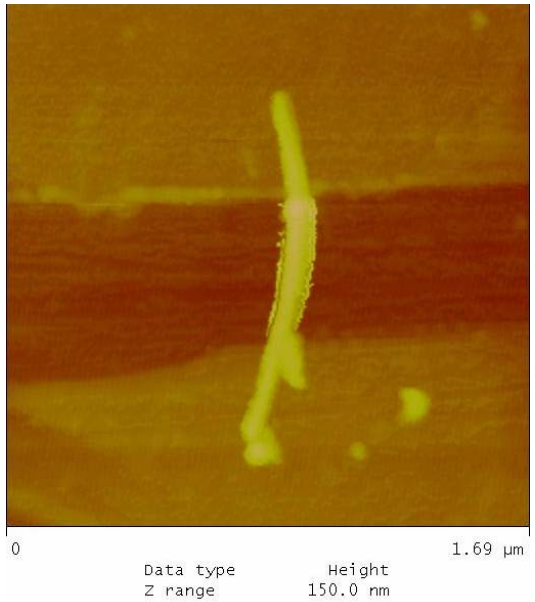
Testing and Results



I-V Characteristic of a CNT

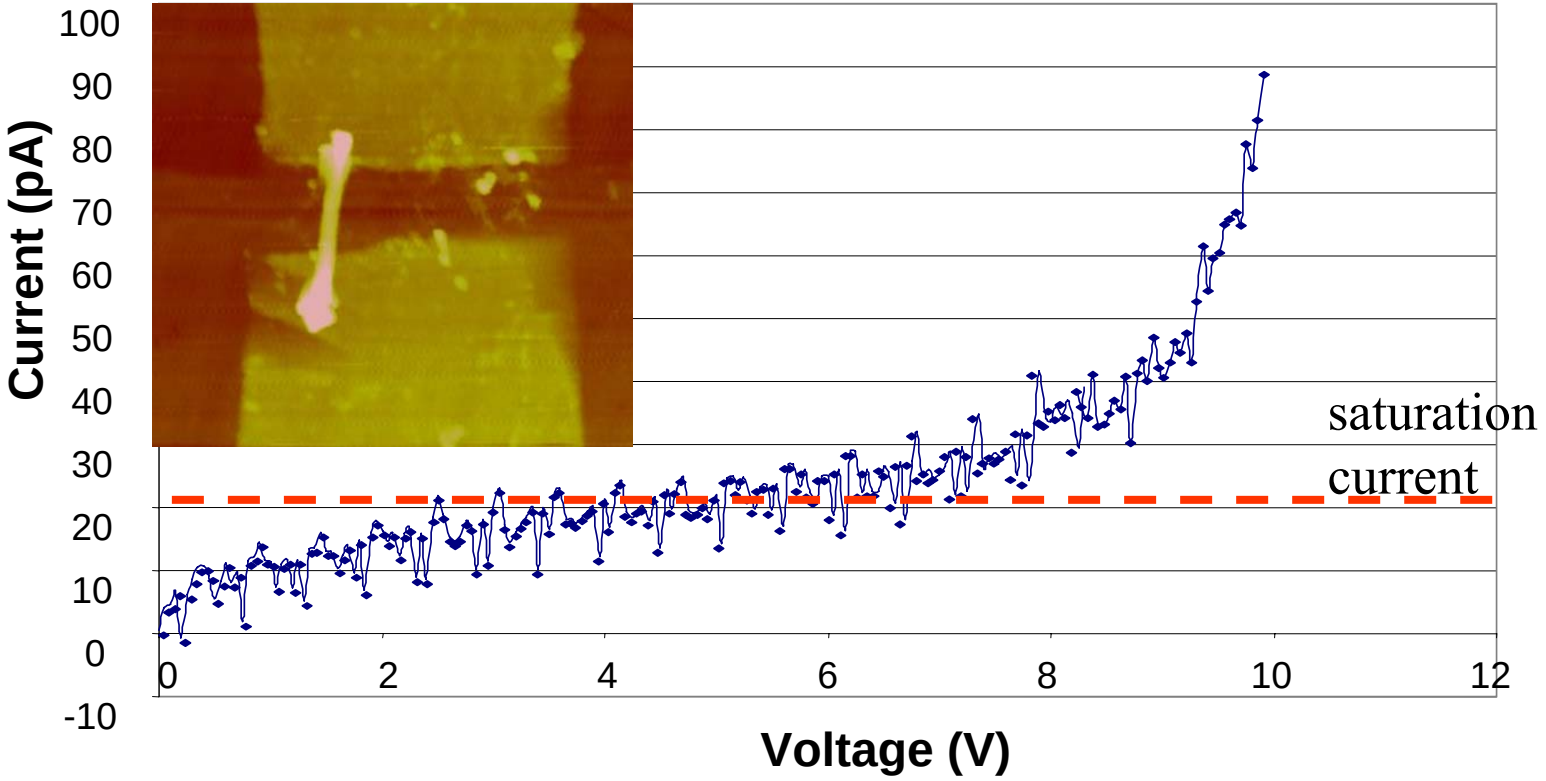
Metallic behavior of CNT, the resistance of CNT is very small. The figure shows the CNT in serial connection with 10M Ohm resistor





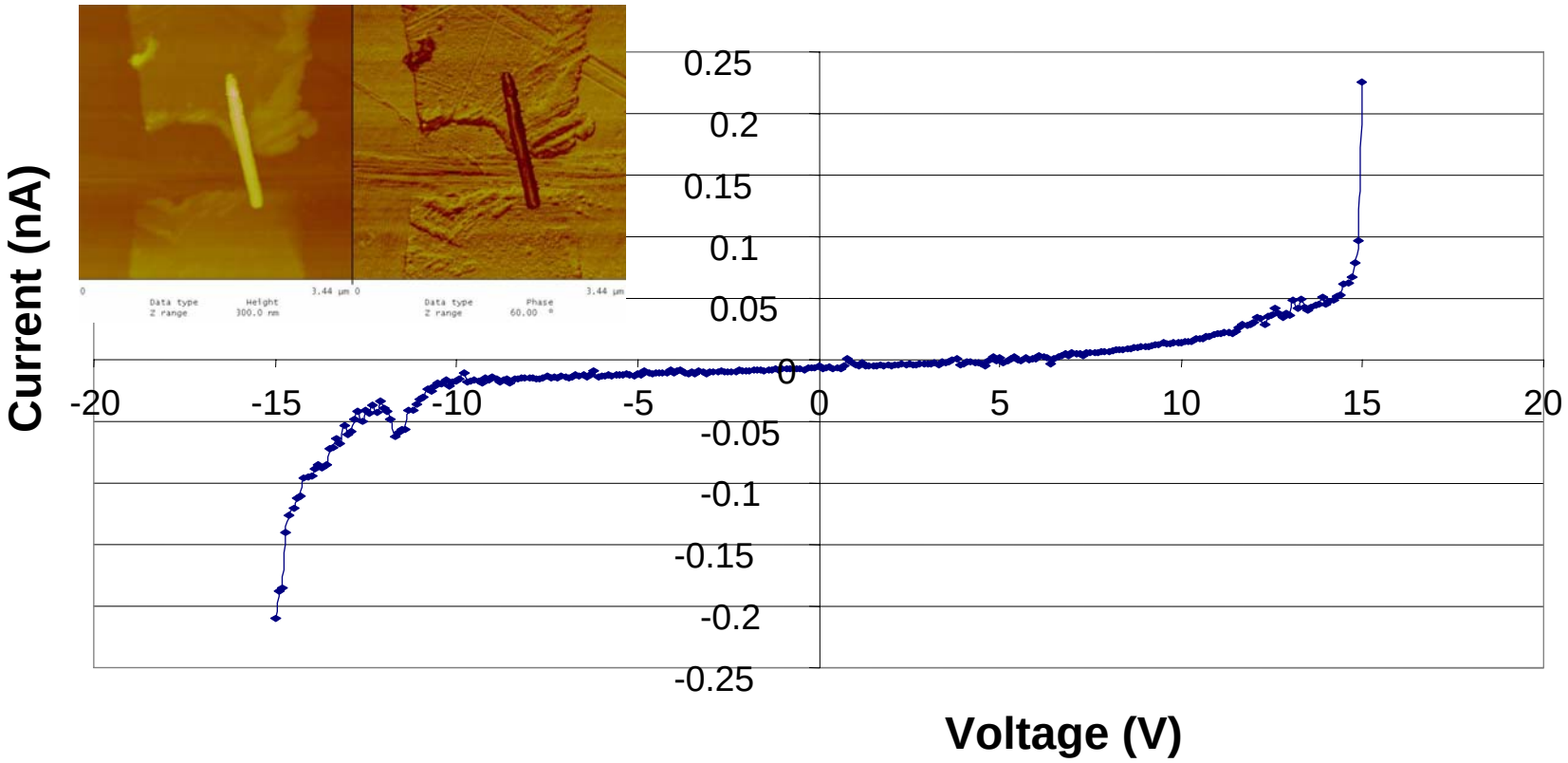
I-V Characteristic of a CNT

Semi-conducting behavior of CNT. The figure shows the CNT in serial connection with 10M Ohm resistor



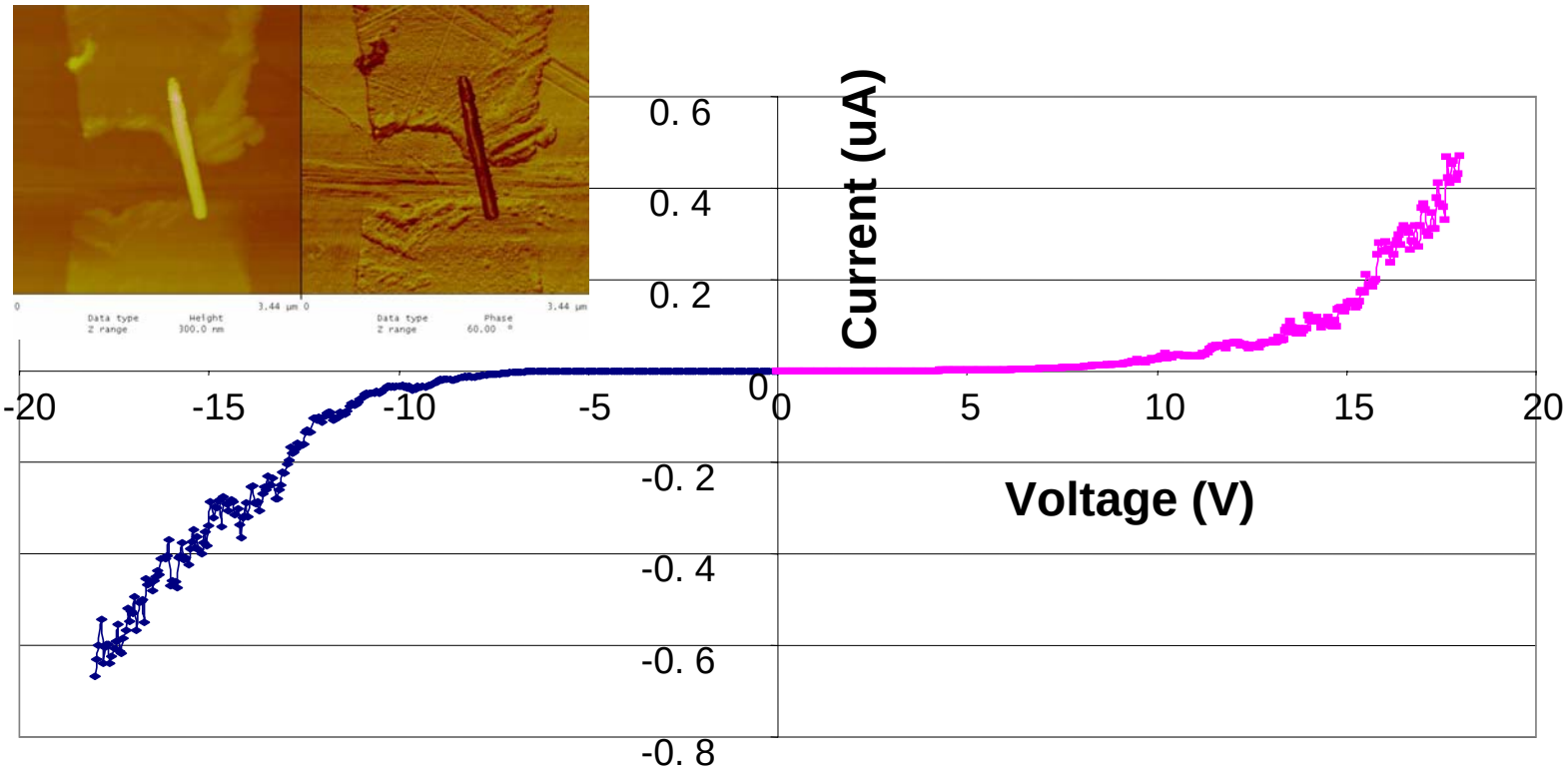
I-V Characteristic of a CNT

Semi-conducting behavior of CNT. The figure shows the CNT in serial connection with 10M Ohm resistor (-15V~15V)



I-V Characteristic of CNT

Semi-conducting behavior of CNT. The figure shows the CNT in serial connection with 10M Ohm resistor. (-18V~18V)



Theory of Brain Info Theory

- Shannon info theory borrowed from physics, Boltzmann statistical mechanics
- Shannon theory is valid for a closed equilibrium system, e.g. Bell Telephone.
- Brain is open dynamic system that has a constant temperature & deals with pairs of sensory inputs 5x2=10 dimensional vector time series, overly redundant data

New Society called "Information Acquisition"

launches by 3 continents 8 countries on June 27-July 3 2005 Hong Kong,
Journal support by WSP, CAS, Japan, Korea, Canada, German, USA

"Intelligence to Machine, Freedom to Mankind"

Brain-central rather PC-central Info Sci & Tech revolution

- Major Force: Next Gen Internets; 3rd Gen Cellular Phones
- Omnipresence—all the time, everywhere
- Launch by 2008 Olympic Beijing, 2010 Worlds Fair Shanghai
- Broad Band requires new array of Transceivers
- new RF towers, similar to HDTV
- Intelligent Interface, user-friendly voice Input
- Innate Natural Intelligence, Voice Video I/O, Robotic Team
- Office Mates, Smart Search Knowledge Engine

Human Civilization Overview
Info Acquisition Perspective

- Historically, the Greeks built theaters everywhere they went, spread art; the Romans built roads everywhere they conquered, spread commerce.
- Much beyond ***(1) roads***, the Romans rebuilt the Greek metropolises with (2) ***water supply pipes*** and (3) ***waste sewage lines.***
- Over a thousand years, growing cities prompted modern western industrialization by means of (4) ***steam-engine trains and railroads***, which accelerated (5) ***ship-line and airline*** commerce that shrink the scale of the earth.
- Recently, info tech launched knowledge economics with the advent of (6) ***Internet***, making last-mile-mobile-broad-band-communications the major challenge of this Century.
- We think the always-on-every-where legacy (7) ***electrical power-line*** could play important role? surveillance of environmental condition need.
- Next Gen Internet (IPv8) based on ***decimal telephone address*** as individual web address could provide a unique and abundant ID for global partnership. Besides, we must solve robust ID ={who,where,when}, one-hook deep web mining for authenticity, pair-wise privacy, and reliability of web knowledge.

High dimensional data representation

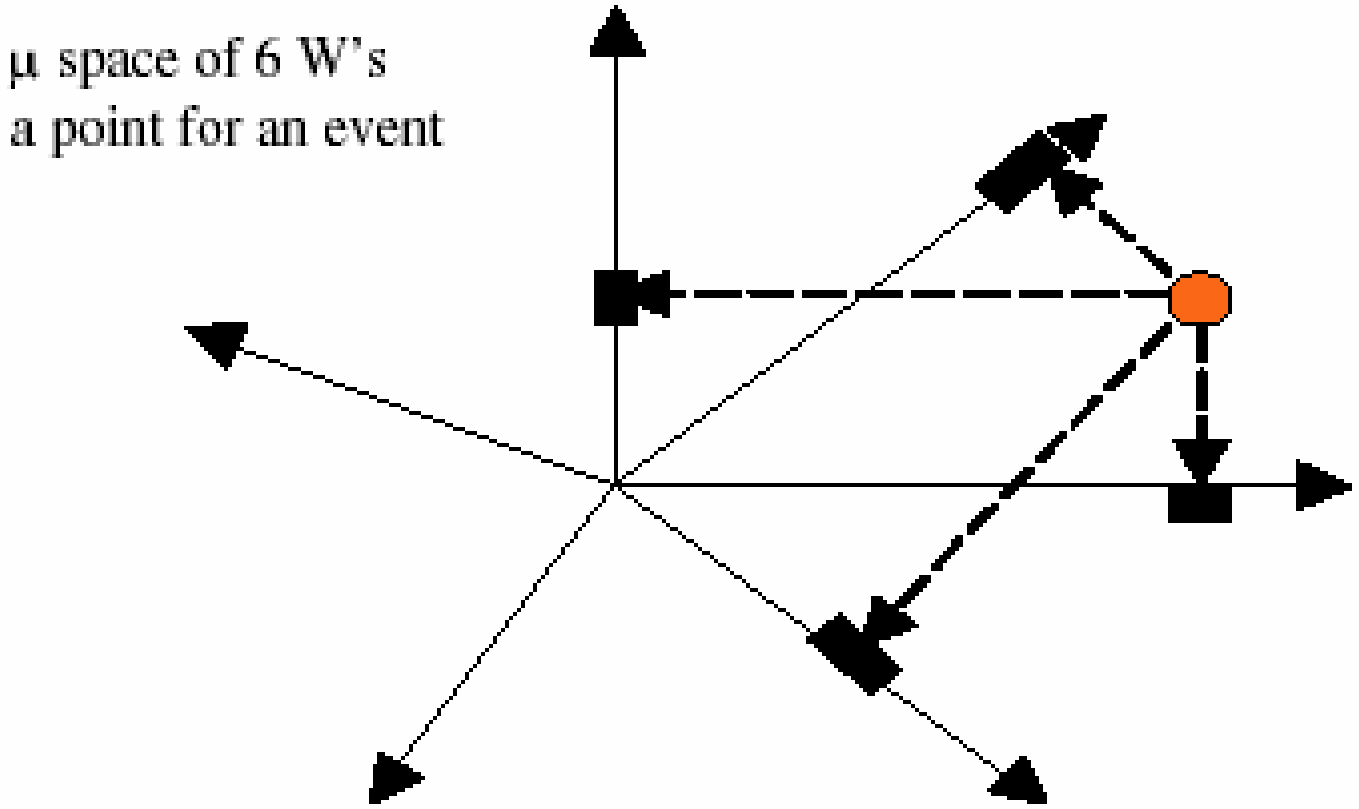


Fig. 1. Data acquisition layer of 6-D μ-space. A point is an event or story in the 6W's "who, when, where, what, how, and why" μ-space; a rough fuzzy set associates a range of projection value indicated by thick line segment.

From 6W data fusion to feature extraction

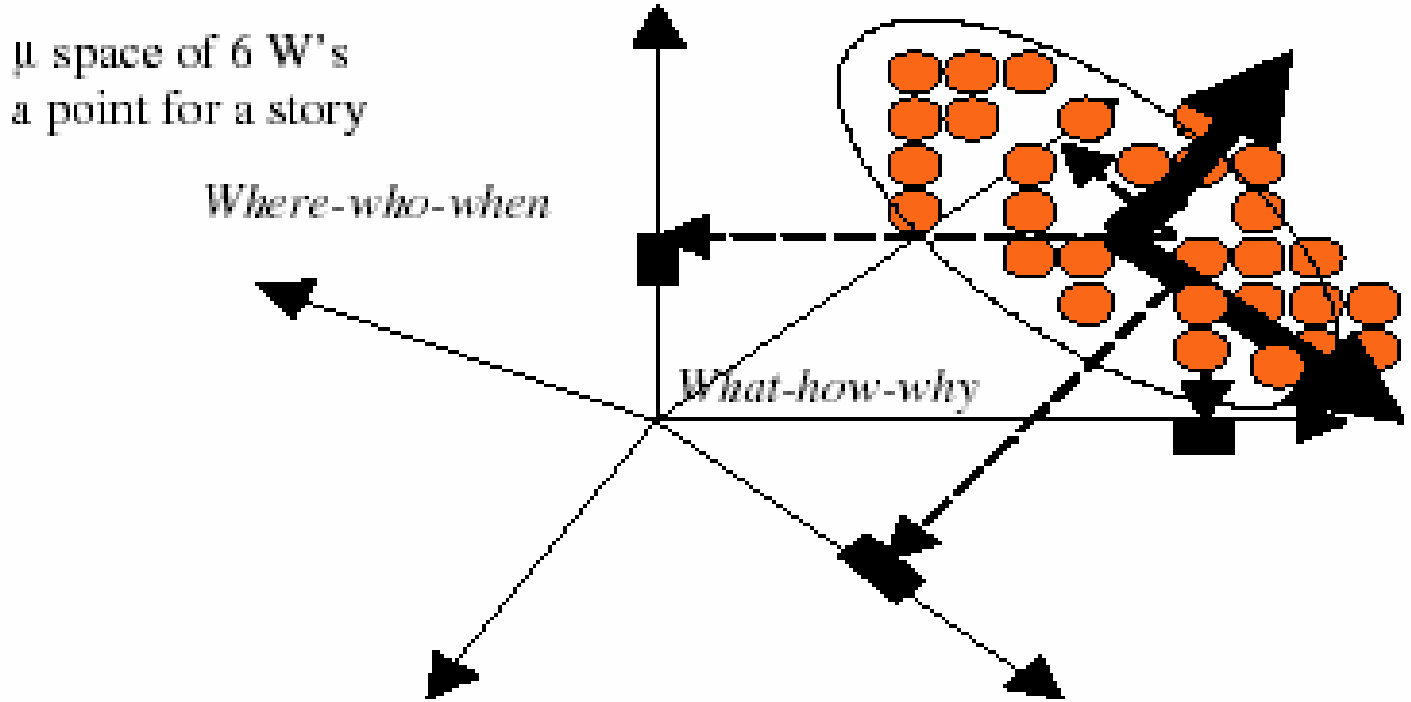
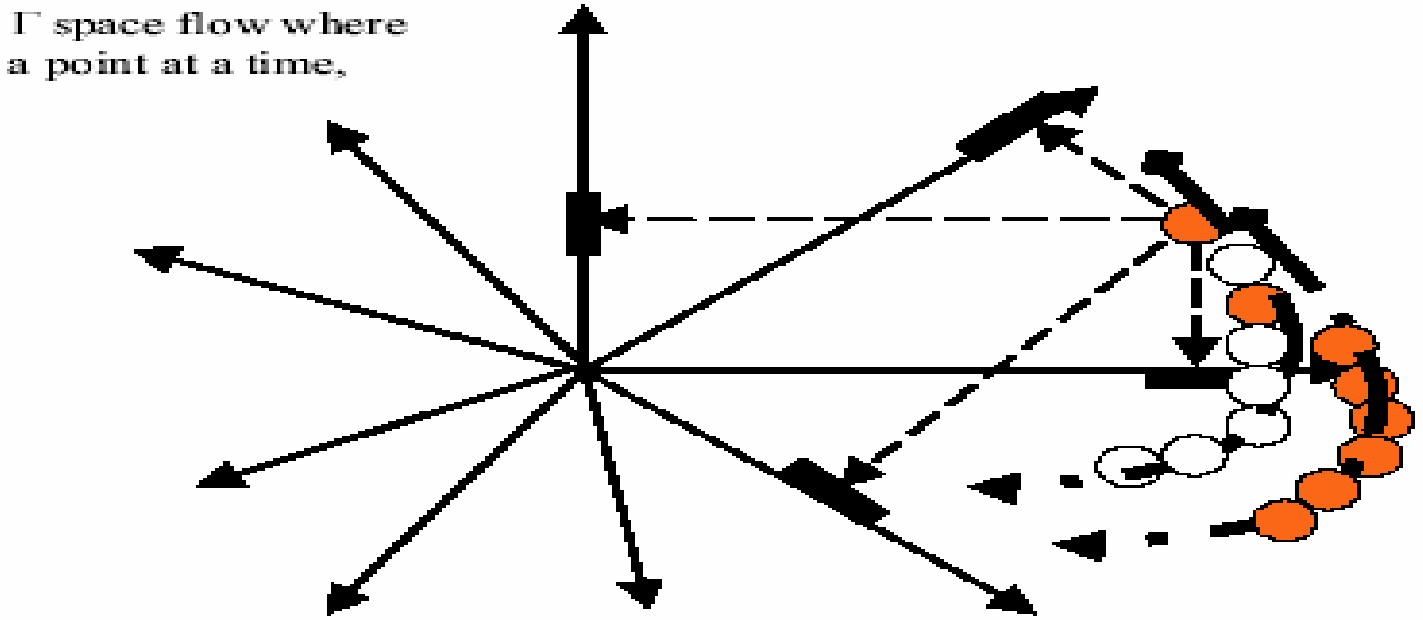


Fig. 2. N newspapers of a daily story yield N dots on 6-D μ-space, where the invariant feature subspace is derived by PCA or ICA as the concatenated space-time event (where-who-when) and cause-effect (what-how-why).

From invariant feature 6N+time to knowledge

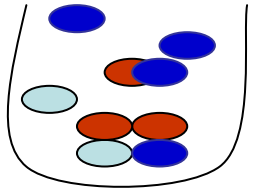


Invariant knowledge bifurcation

Fig. 3. A time-invariant feature space becoming known as the knowledge space. In this example, "U.S. President John F. Kennedy was assassinated at Dallas, Texas on November 22, 1963" two major-track bifurcations might occur in time in the subspace of 6ND Γ-space of N reports over years those major and minor feature axes in the 6D μ-space. The knowledge discovery can be "a track of conspiracy theory and a track of loner lunatic theory," and one becomes dominating over the other in time as the evidence is presented.

Shannon's Theory came from a closed equilibrium

Independent & identical color balls ***(N=R+G+B)*** of single bucket have the ***a priori*** combinatorial state space ***W=N!/(R!G!B!)***



Boltzmann's Tomb Stone $S = K_B\, LogW$

of which Shannon formula follows

$$S = -\,NK_B\, \Sigma'_i\, s_i \log s_i > 0$$

where the prime indicates $\Sigma_i\, s_i = 1$, the minus sign for $\log s_i < 0$ as $s_i < 1$.

Proof: ***If N=R+G+B, then 1=R/N+G/N+B/N and use is made of Stirling formula:*** $\log N! = \log N(N-1)(N-2)\ldots = N \log N - N$

Then,

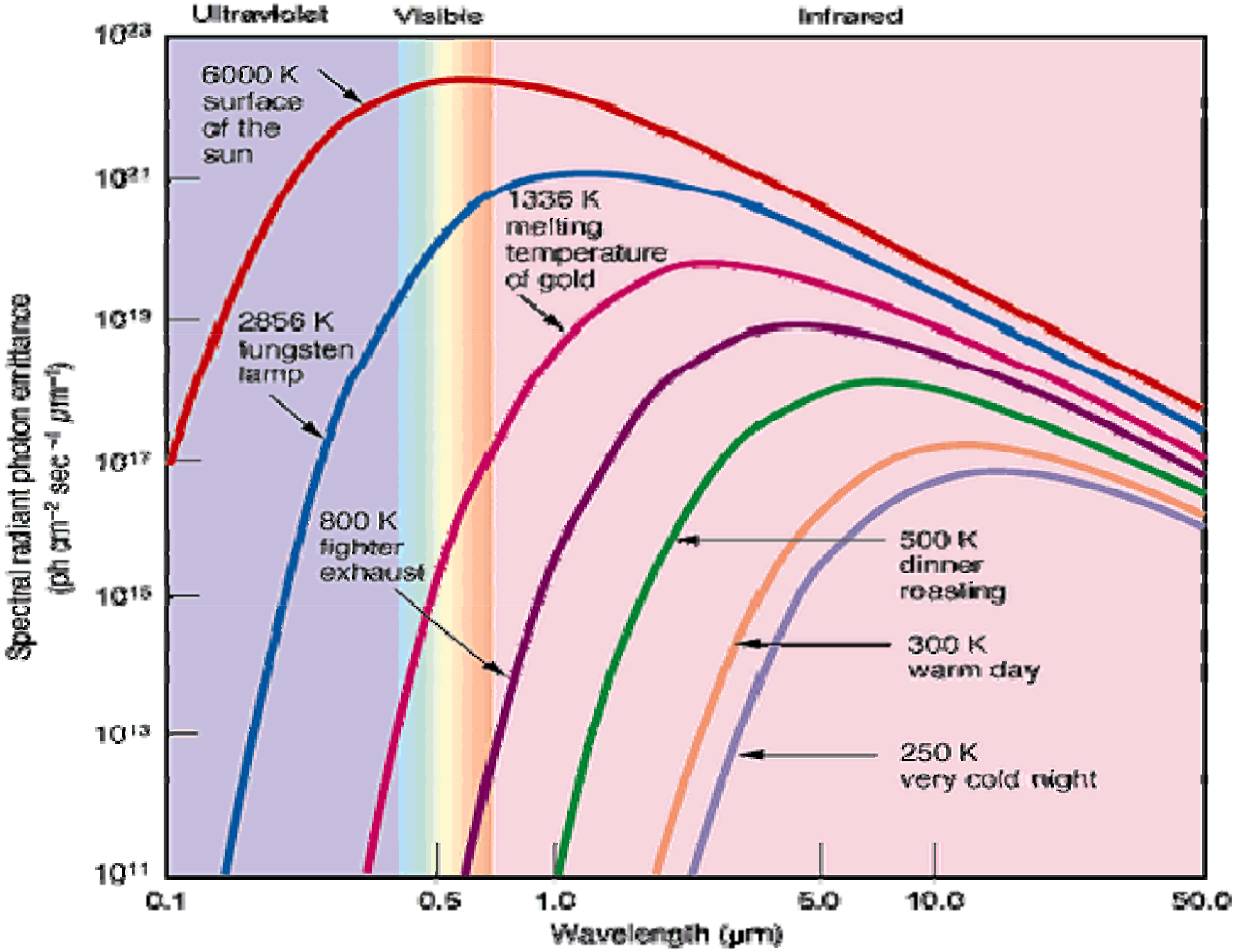
$$S/N = K_B\, Log(\, N!/R!B!G!)$$

$$= K_B[(R+G+B)\log N - \underline{N} - (R\log R - \underline{R} + B\log B - \underline{B} + G\log G - \underline{G})]$$

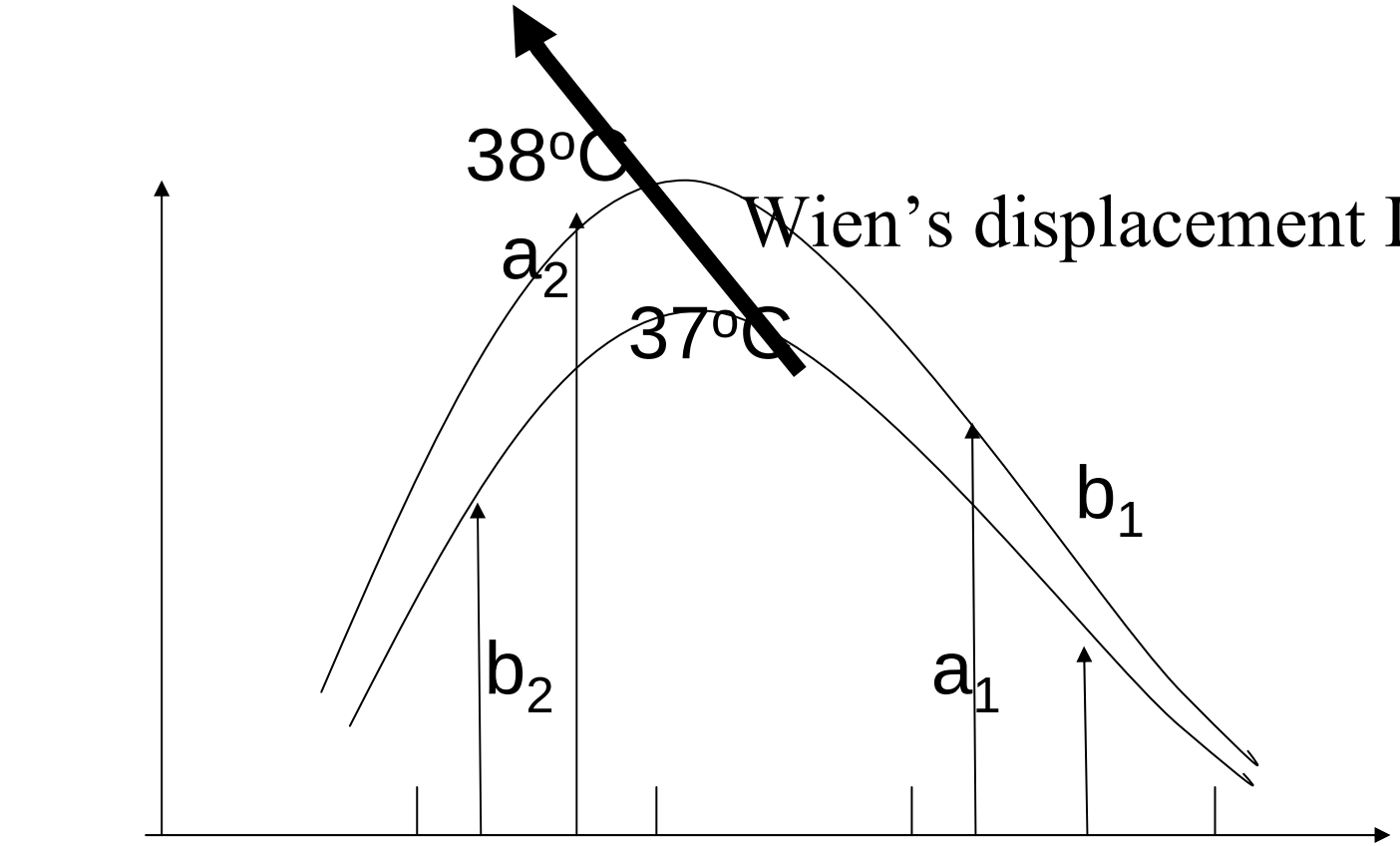
$$= K_B\,[-R/\log(R/N) - B\,\log(B/N) - G\,\log(G/N)]$$

$$= -\,NK_B\,[s_1 \log s_1 + s_2 \log s_2 + s_3 \log s_3] = -\,NK_B\,\Sigma'_i\, s_i \log s_i .$$

QED

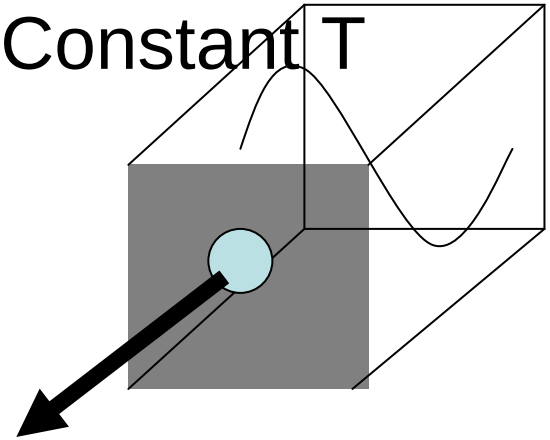

Ultraviolet
Visible
Infrared
6000 K surface of the sun
2856 K tungsten lamp
1335 K melting temperature of gold
800 K fighter exhaust
500 K dinner roasting
300 K warm day
250 K very cold night
Spectral radiant photon emittance (ph m⁻² sec⁻¹ μm⁻¹)
10²³
10²¹
10¹⁹
10¹⁷
10¹⁵
10¹³
10¹¹
0.1
0.5
1.0
5.0
10.0
50.0
Wavelength (μm)

Planck Radiation



Constant T

Blackbody Expt.



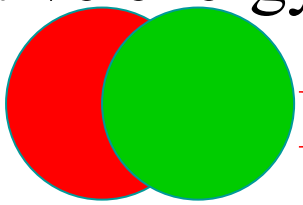
Mathematical Ill-posed Inverse Problem "Power of Pairs"

give vector time series X(t) having numerous feature S(t) decompositions upon which one chooses the most probable equilibrium answer imposed by min.Helmholtz free energy

Guess what were hidden real positive energy sources 3 &5 ?

e.g. $2\text{x}5 + 1\text{x}3 = 13$

$1\text{x}5 + 3\text{x}3 = 14$



Power of Pairs

Given two resulting numbers 13 & 14 ***Given data X***,

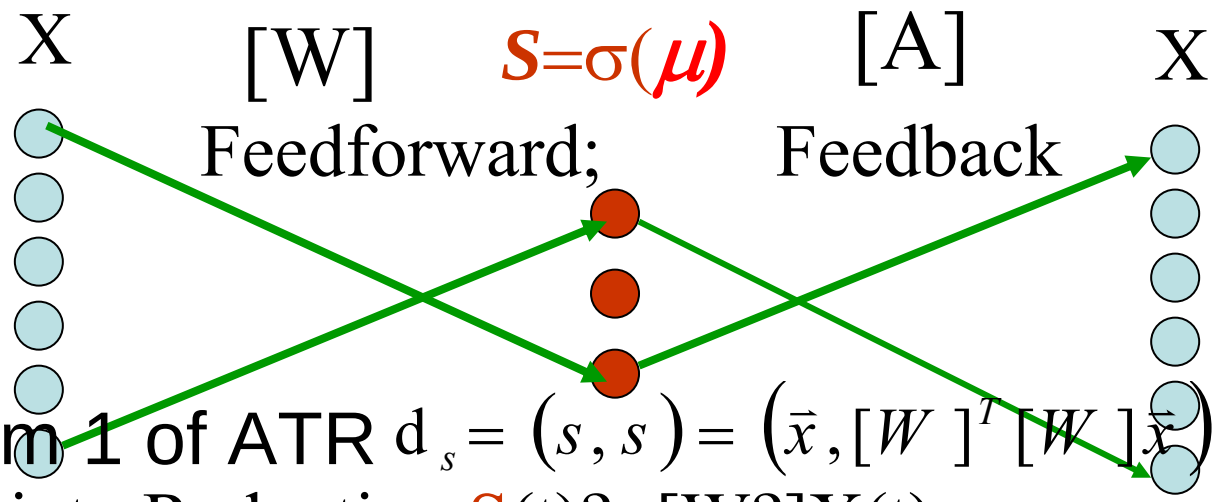
find both unknown mixing matrix & sources [A?]S?

$$\begin{bmatrix}13\\14\end{bmatrix} = 5\begin{pmatrix}2\\1\end{pmatrix} + 3\begin{pmatrix}1\\3\end{pmatrix} = \begin{bmatrix}2 & 1\\1 & 3\end{bmatrix}\begin{pmatrix}5\\3\end{pmatrix}; \vec{X} = [A]\vec{S} = s_1\vec{a} + s_2\vec{b};$$

we can always normalize the data X

for unknown unit feature vectors $\vec{a}, \vec{b}$

Unsupervised Learning Based on Thermodynamic Equilibrium by Min. Free Energy H=E-TS; E=μ([W]X-S)= λ(X-[A]S) 1st order LMS error energy



Theorem 1 of ATR $d_s = (s, s) = (\vec{x}, [W]^T [W] \vec{x}) = d_{\vec{x}}$

- Uncertainty Reduction S(t)?=[W?]X(t);
- Associative recall X(t)=[A?]S(t)?
- Theorem 2 Unsupervised Hebb info distance

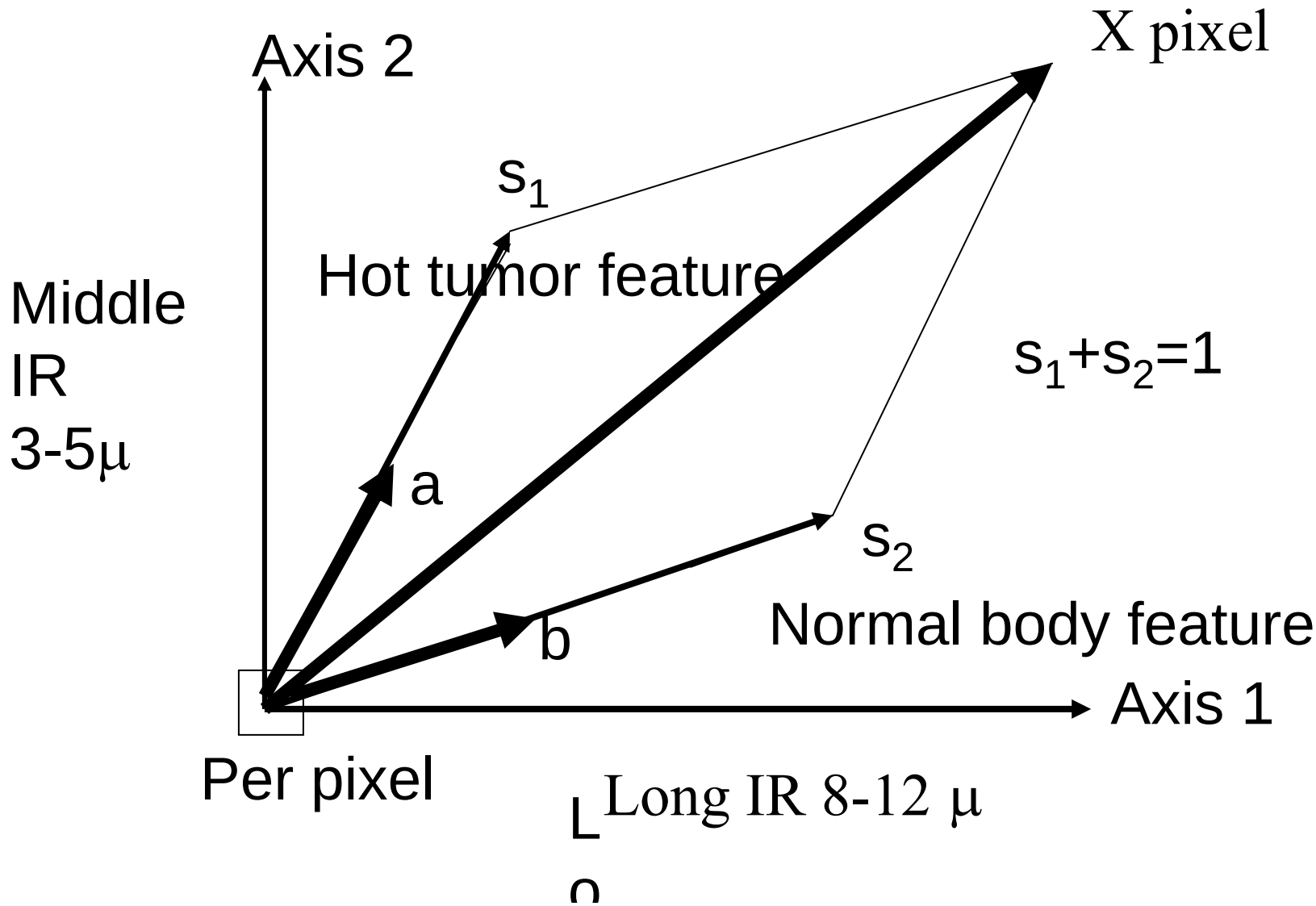
full rank learning $(d[W]/dt) = -(dH/d[W])[W^T][W] = \mu X [W^T][W]$

- Theorem 3 Generalized Equal Partition Sigmoid logic

$$\sigma(\mu) = S = [W]X$$

knowing component s_1, s_2 *follows*

uniquely vector features $[\vec{a}\ \vec{b}\ ] \equiv [\ A\ ]$

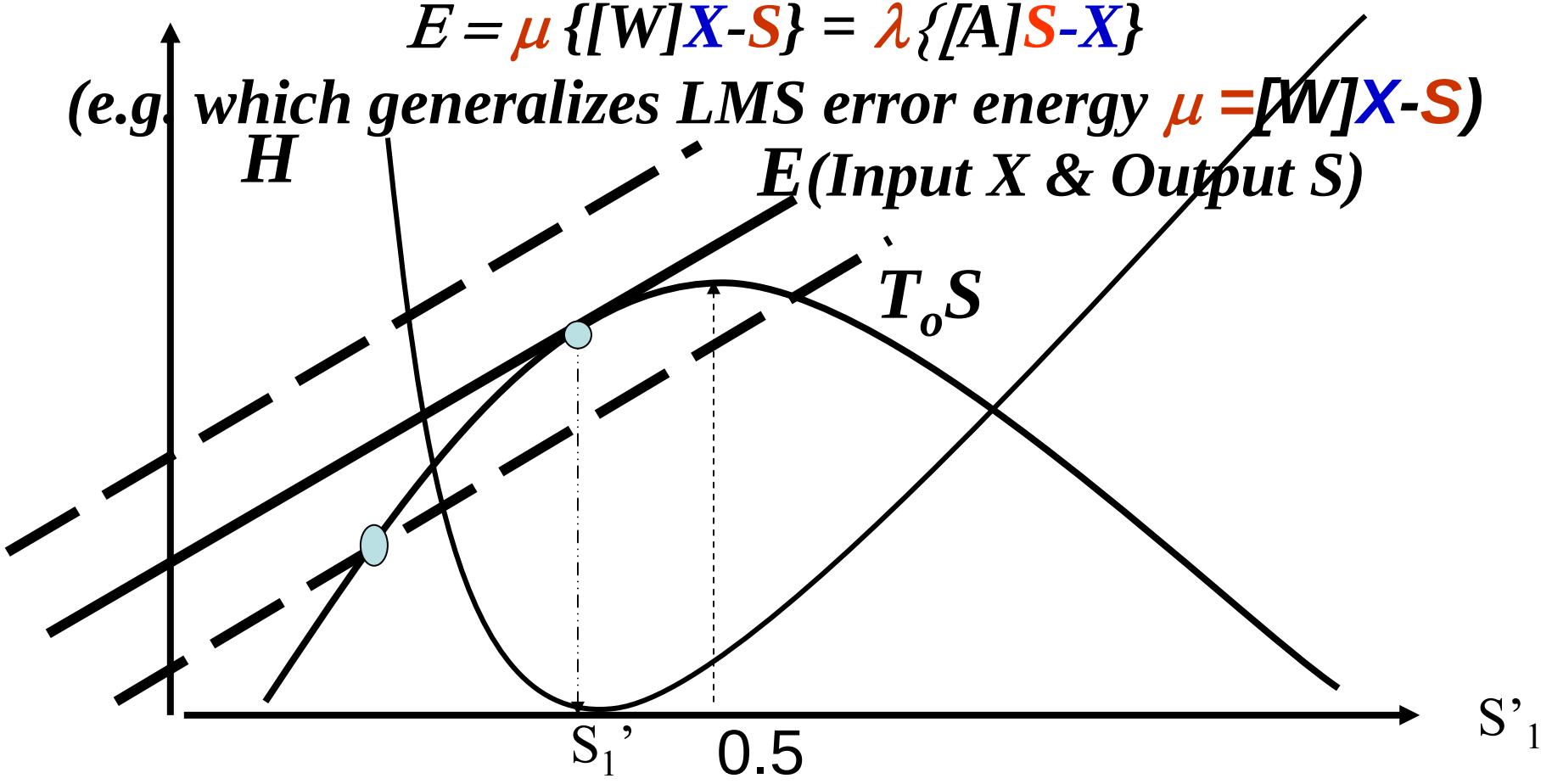


Graphical Proof of Uniqueness of sources

We postulate a linear information I/O energy

$$E = \mu\{[W]X - S\} = \lambda\{[A]S - X\}$$

(e.g. which generalizes LMS error energy μ =[W]X-S)



Local Maximum Entropy for Most Probable Solution

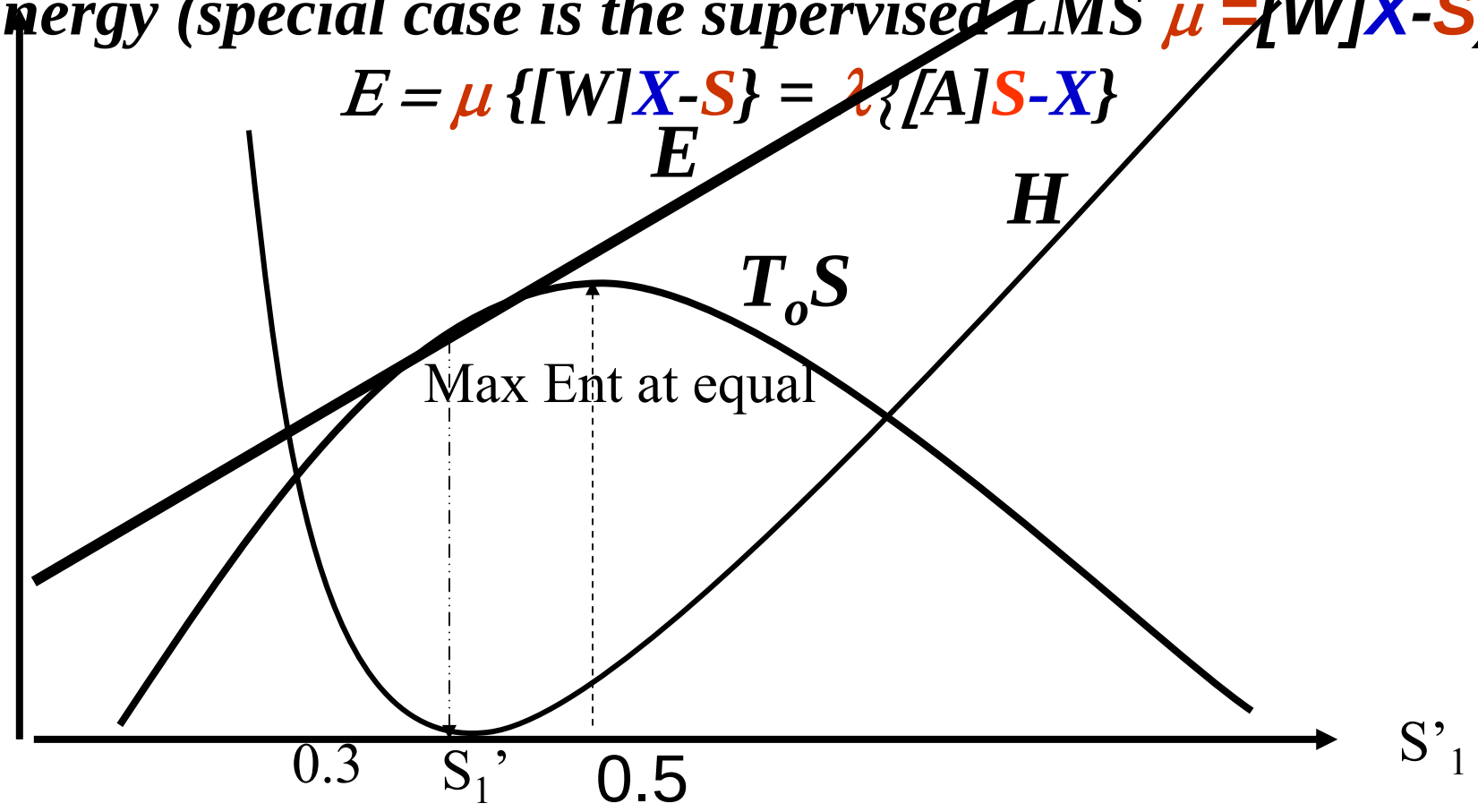
$$s_2' = 1 - s_1'$$

theory

to open systems with I/O energy E with unique inverse solutions

Unsupervised Lagrange Constraint Information I/O Energy (special case is the supervised LMS μ =[W]X-S)

$$E = \mu\{[W]X-S\} = \lambda\{[A]S-X\}$$



Local Maximum Entropy for Most Probable Solution

s_2'=1-s_1'

ANN involves dendrite signal pre-conditioning

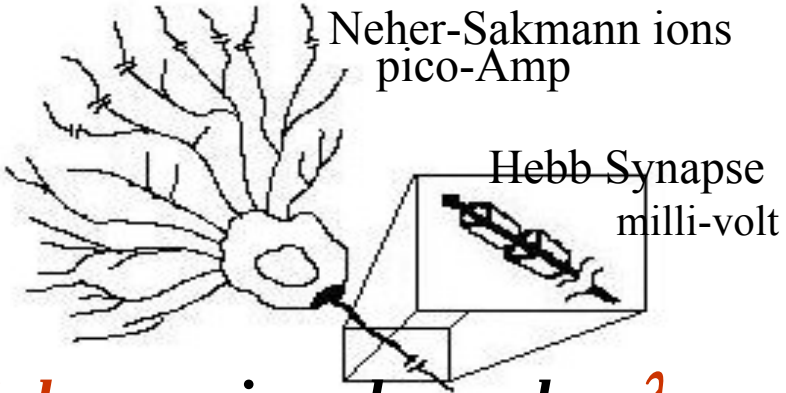
Isothermal Free Energy $H = E - T_o S$ Helmholtz

"Gasoline does PdV work after the heat waste"

Where a homeostasis body T_o=37°C,

$$K_B T_{room} = \frac{1}{40} eV$$

Entropy $S = - N K_B \Sigma_i s'_i \log s'_i$



Local Lagrange Constraint, 1991 Neher-Sakmann ion channels μ, λ.

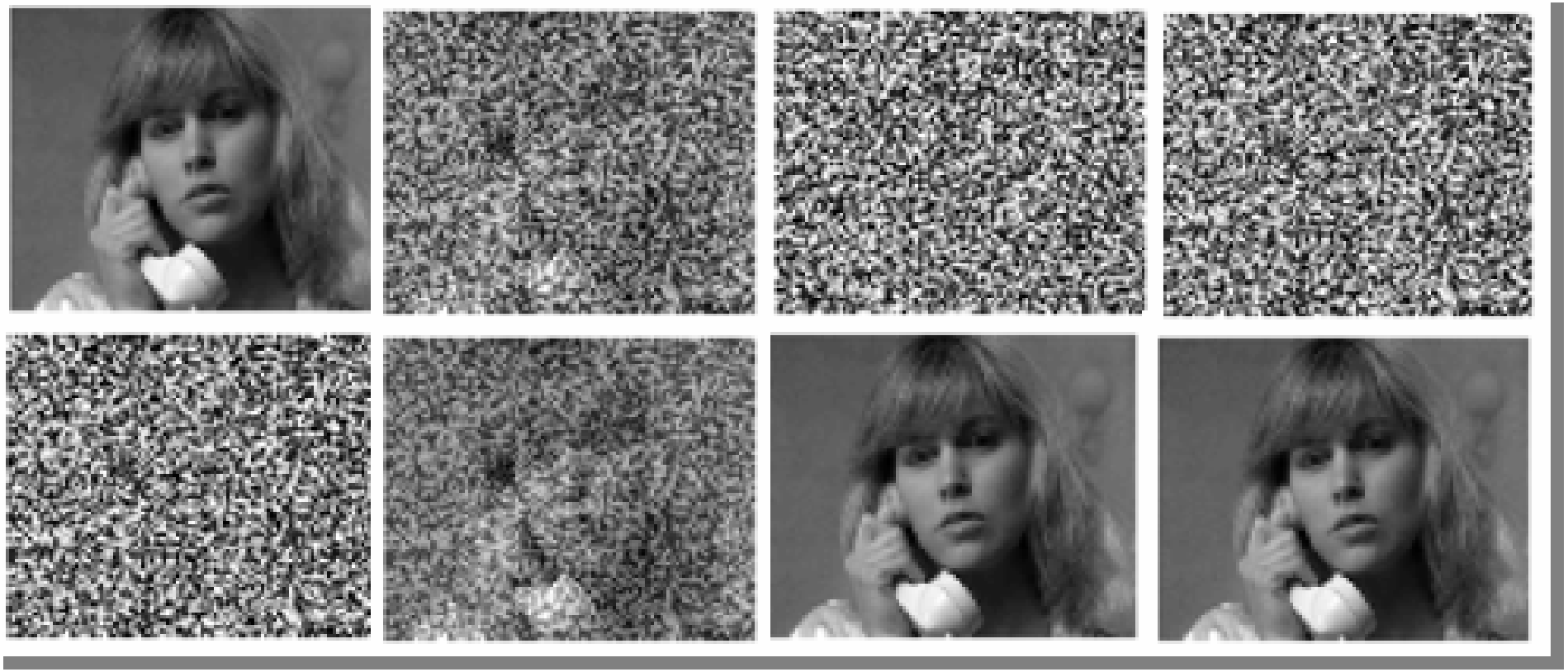
AM uses 10,000 neurons

$$E = \vec{\mu}_{10^{-12}\ Amp} \left\{ [W]_{10^{-3}\ Volt} \vec{X} - \vec{S} \right\} = \vec{\lambda} \left\{ [A] \vec{S} - \vec{X} \right\}$$

Lagrange multipliers were conjectured to play the role of ion current & housekeeping glial cells in neurobiology for the unsupervised internal tutor.

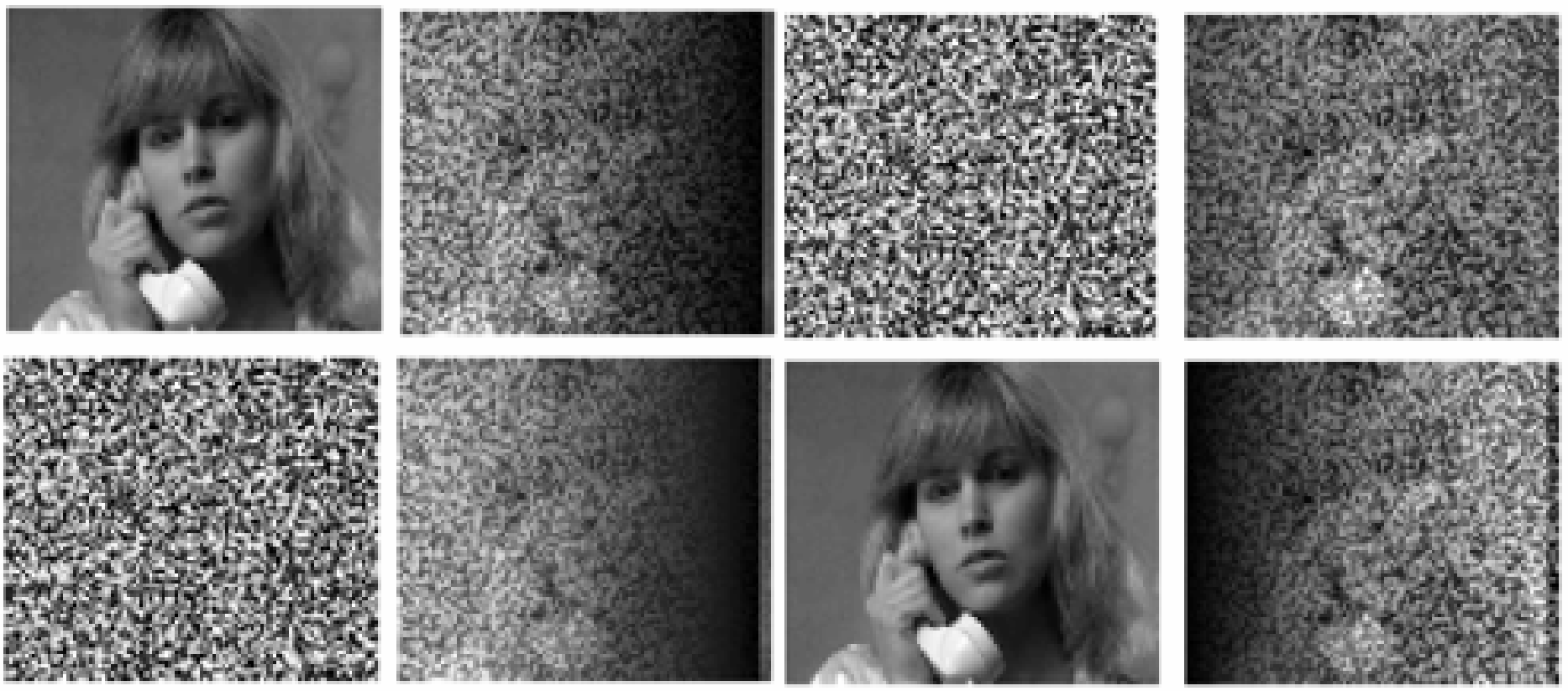
No difference Space Invariant Imaging, single [A]

From left to right: (i) source images;
(ii) linear space-invariant mixtures;
(iii) recovered images using **linear LCNN** algorithm
(iv) recovered images using BSAO **Infomax ICA** algorithm



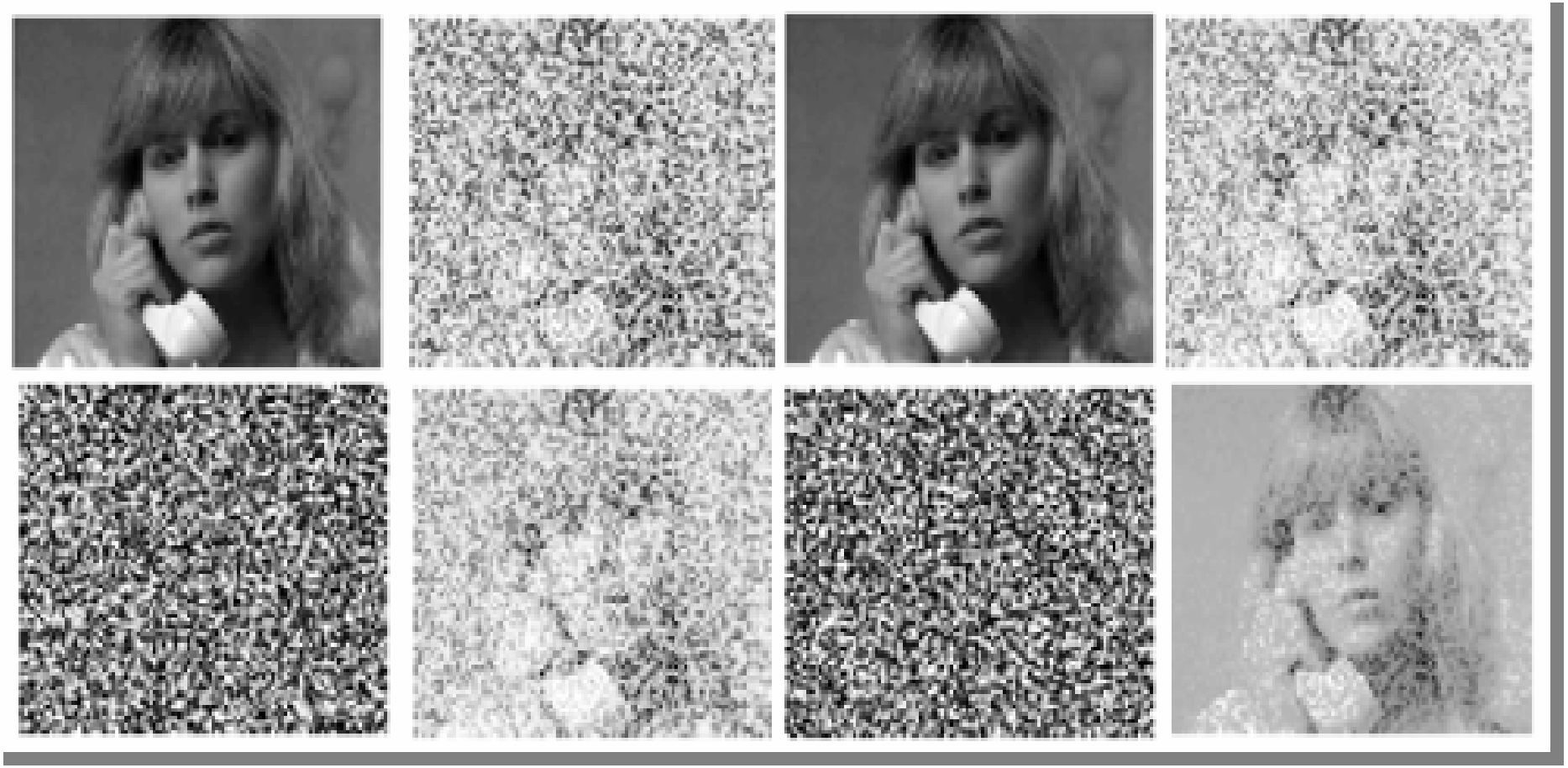
Space Variant Imaging, variant [A]

From left to right: (i) source images;
(ii) linear space-variant mixtures;
(iii) recovered images using **LCNN** algorithm
(iv) recovered images using **Infomax ICA** algorithm



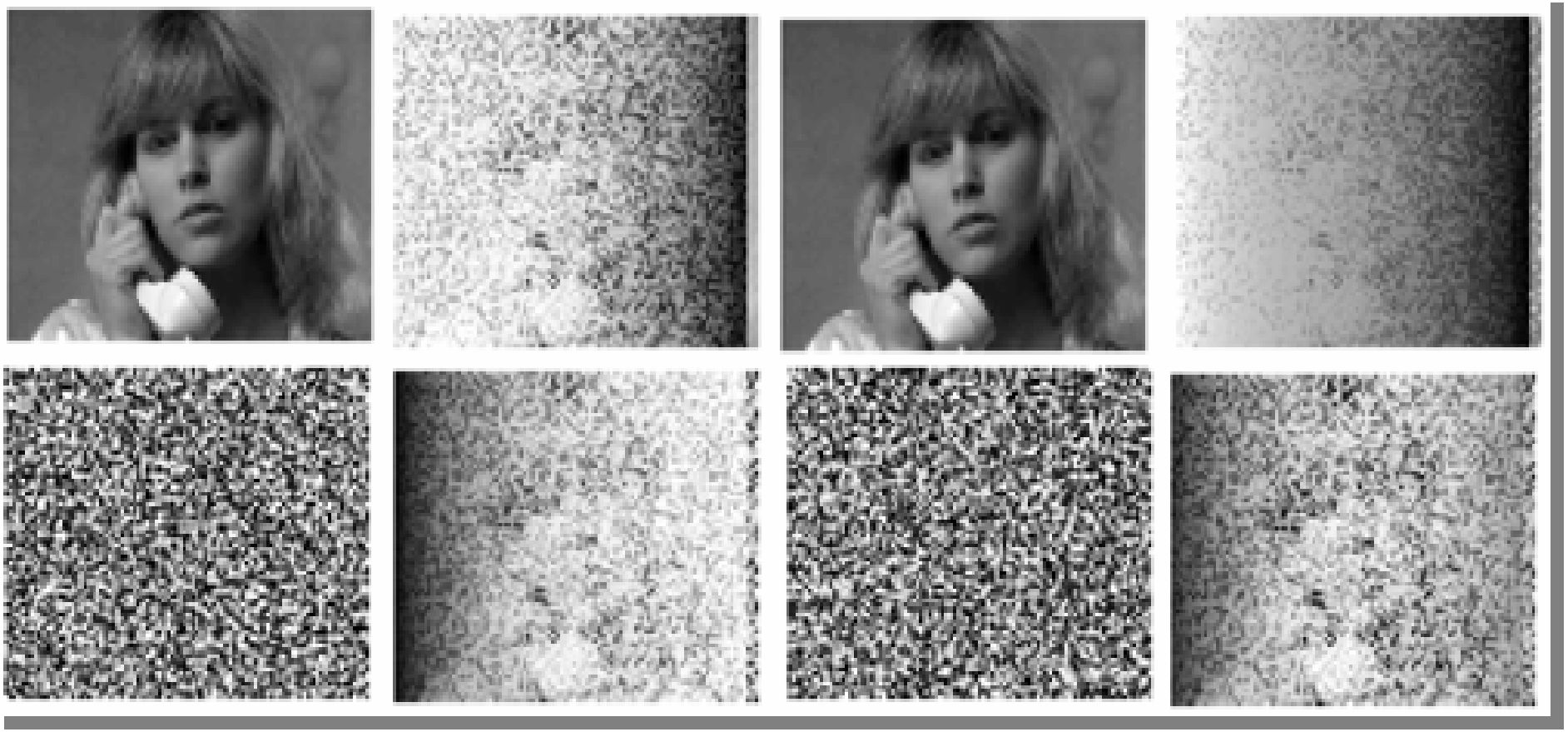
Nonlinear Space Invariant Imaging, single[A]

From left to right: (i) source images;
(ii) nonlinear space-invariant mixtures;
(iii) recovered images using **nonlinear LCNN** algorithm
(iv) recovered images using **Infomax ICA** algorithm



Nonlinear Space-variant Imaging, variant [A]

From left to right: (i) source images;
(ii) nonlinear space-variant mixtures;
(iii) recovered images using nonlinear LCNN algorithm
(iv) recovered images using Infomax algorithm



Closed Equilibrium Theorem of Maximum A Priori Entropy:
m-Equal-Partition Law

Given a-priori entropy of m independent identical parts

$$S = -NK_B \Sigma'_{i=1,m} s_i \log s_i$$

$$= -NK_B \Sigma_{i=1,m} s_i \log s_i + NK_B (\mu_o+1)(\Sigma_{i=1,m} s_i - 1)$$

Proof: *One can not conduct partial differentiation as all components are coupled by the unit norm, unless we remove it by Lagrange multiplier* (μ_o+1)

$$S = -NK_B \sum_{i=1,m} s_i \log s_i + NK_B (\mu_o+1)(\sum_{i=1,m} s_i - 1)$$

$$\frac{dS}{ds_j} = \sum_k \frac{\partial S}{\partial s_k}\frac{ds_k}{ds_j} \to \frac{\partial S}{\partial s_j}_{const.k} = NK_B(\log s_j + 1) + NK_B(\mu_o+1) = 0$$

$s_j = \exp(\mu_o)$ *to determine Lagrange multiplier by the constraint value*

$$\sum_{j=1,m} s_j = \sum_{j=1,m} \exp(\mu_o) = m\exp(\mu_o) = 1$$

$$s_j = \exp(\mu_o) = \frac{1}{m}$$

Q.E.D.

Open dynamic equilibrium by the minimum of Helmholtz free energy $H=E-T_0S$ generalized the equal-partition law to Sigmoid Law

$$H([W],\vec{S}) = E - T_0 S$$

$$= \vec{\mu}^T \left[[W]\vec{X} - N\vec{S} \right] + NK_B T_0 \sum_{i=1}^{m} s_i \log s_i - NK_B T_0 (\mu_0 + 1)\left(\sum_{i=1}^{m} s_i - 1 \right);$$

$$N = Norm\ \vec{X}$$

$$Equilibrium\ at\ \frac{\partial H}{\partial s_j} = 0 \quad \Rightarrow$$

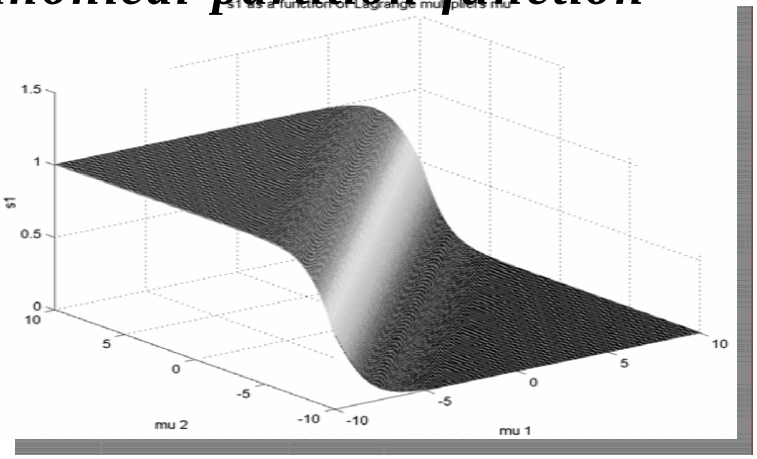
$$\vec{s}_j == exp\left(\frac{\mu_j}{K_B T_0} + \mu_0 \right) = \frac{1}{1 + \sum_{\substack{i=1 \\ i \neq j}}^{m} exp\left(\frac{1}{K_B T_0}(\mu_i - \mu_j) \right)} = \sigma(\vec{\mu}_j)$$

$$where\ use\ is\ made\ of\ the\ constraint\ value\ \sum_{j=1}^{m} s_j = 1$$

$$\Rightarrow\ exp(-\mu_0) = \sum_{j=1}^{m} exp\left(\frac{\mu_j}{K_B T_0} \right) \equiv Z\ Canonical\ partition\ function$$

One can verify the free energy as displacement with constraint force

$$\frac{\partial \mu_0}{\partial \mu_j} = -s_j$$



Real World Unsupervised Applications:

$$X_{pixel}=[A_{pixel}?]S_{pixel}?$$

(1)***satellite remote sensing without the ground truth***
(2) **early pre-cancer tumor detection without library features**
(3) ***animal unsupervised learning artificial neural nets***.

All require a ***single pixel super-resolution*** (i.e. by nature of a large pixel footprint about 30mx30m and earliest warning of malign tumor about sub-mini-meter) where the unknown heat transport mixing matrix [A] (unsupervised features characterizing space-variant point spread function or impulse response function of the unknown turbulent air propagation or breast tissue) that varies from pixel to pixel.

Unsupervised ANN

We derive ANN from top-down & bottom-up observations:

(1) Bottom-up:vector time series input from pairs sensors

(2) Top-down: isothermal equilibrium condition.

We obtain three important results as follows:

(**i**))We shall first review Boltzmann entropy that leads to Shannon formula by Stirling then generalize it.

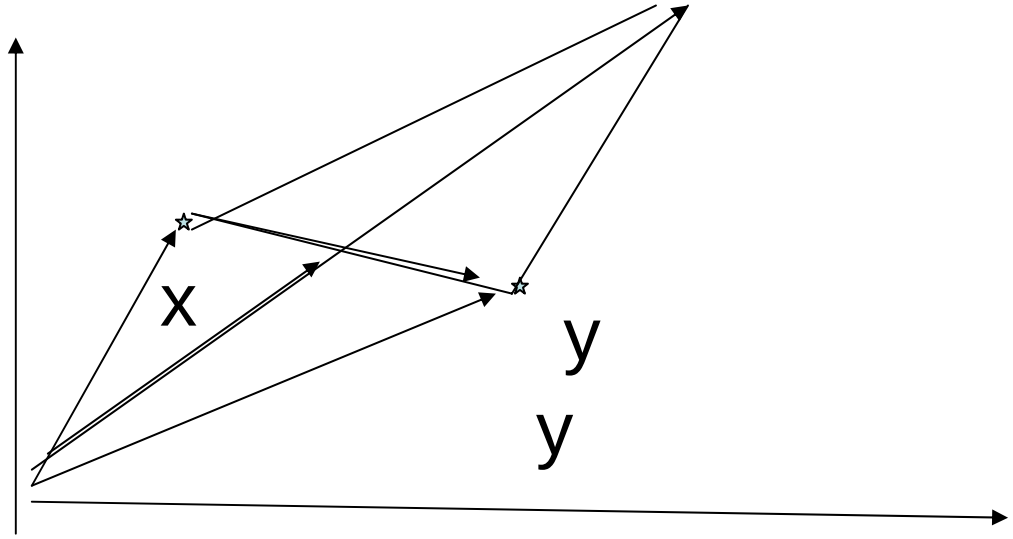
(ii) The equilibrium equal partition law, which becomes for brain open equilibrium to be sigmoid logic without assuming it.

(iii) We assume Lagrange Constraint m to be Neher-Sakmann pico-Ampere ion channels 1991.

(iv) We derive unsupervised Hebb rule from free energy min.

(v) Constraint optimization based on Cauchy fast simulated annealing at cooling schedule $T=T_o/(1+t)$, $t>0$.(vs.T_o/log t)

Self Organization Map of Kohonen



$X_{new} = (X_{old}+Y)/2 = X_{old} + (Y - X_{old})/2$

$\langle Xn \rangle = \langle Xn-1 \rangle + (1/n-1)(Xn - \langle Xn-1 \rangle)$

SOM is supervised i.e. a labeled data set centroid feature extraction

Kohonen Self Organization Map Centroid algorithm
Batch Average versus Sequential Average

$$< X >_{m+1} \equiv \frac{1}{(m+1)w} \sum_{i=1}^{m+1} x_i w_i$$

$$= \frac{(m+1-1)}{m+1} \frac{1}{m} \sum_{i=1}^{m} x_i + \frac{1}{m+1} x_{m+1}$$

$$=< X >_m + K\,(x_{m+1} - < X >_m);$$

$$Kalman \quad gain \quad K = \frac{1}{m+1}:$$

$$if \quad uniform \quad weight \quad w_i = w = 1$$

$X(t) = (x(t), x(t-1), \ldots, x(t-m+1))^T$ *Wiener missile guidance AR 1950*

$u_i(t+1)=w_i(t)^T X(t)$ *predicts the future from the past*

$E = \langle [u_i(t+1) - x(t+1)]^2 \rangle$ *LMS*

$v_i(t+1) = \sigma(w_i(t)^T X(t)) \cong w_i(t)^T X(t) = u_i(t+1)$ *Nonlinear ANN*

$dw_i/dt = -\partial E/\partial w_i = 2 \langle [w_i^T X(t) - x(t+1)] X(t)^T \rangle = 0$ *Fixed p.*

$C_s = \langle x(t)\, x(t-s) \rangle$ *stationary Teoplitz matrix*

$$\begin{bmatrix} C_0 & C_1 & C_2 & \cdots & C_{m-1} \\ C_1 & C_0 & C_1 & \cdots & C_{m-2} \\ C_2 & C_1 & C_0 & \cdots & C_{m-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{m-1} & C_{m-2} & \cdots & C_1 & C_o \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_m \end{bmatrix}$$

$dw_i/dt = -\partial E(v_i, x(t+1))/\partial w_i$ Nonstationary *ANN 1970*

$dw_i/dt = \langle \partial K(V)/\partial w_i \rangle = 0;\ K(V) = \langle V^4 \rangle - 3 (\langle V^2 \rangle)^2$

$\langle v_1 v_2 v_3 v_4 \rangle_G =$ **Unsupervised Fast-ICA 2000**

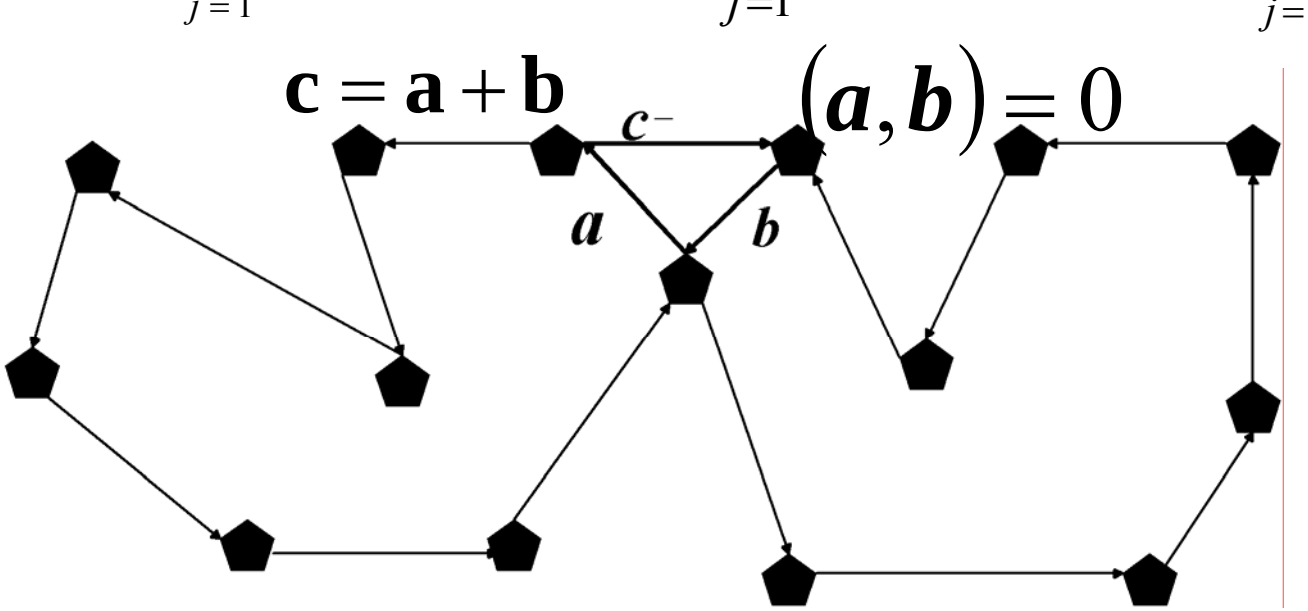
$\langle v_1 v_2 \rangle_G \langle v_3 v_4 \rangle_G + \langle v_1 v_3 \rangle_G \langle v_2 v_4 \rangle_G + \langle v_2 v_3 \rangle_G \langle v_1 v_4 \rangle_G$

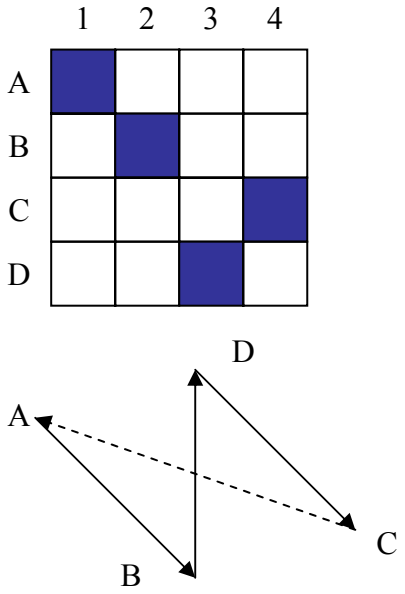
Classical ANN Constraint Optimization by Divide & Conquer Solving N-P Complete TSP

A vector tour sum = minus-one city tour was decomposed into two $\boldsymbol{a}, \boldsymbol{b}$ separate minus-one city tours such that an orthogonal decomposition could be found.

$$D = \sum_{j=1}^{N} \sqrt{(\boldsymbol{c}_j, \boldsymbol{c}_j)} \qquad \sum_{j=1}^{N} \boldsymbol{c}_j = 0 \qquad \mathbf{c} = \sum_{j=1}^{N-1} \mathbf{c}_j = -\mathbf{c}_N$$

$$\mathbf{c} = \mathbf{a} + \mathbf{b} \qquad (\boldsymbol{a}, \boldsymbol{b}) = 0$$





Theorem $\min(c, c) = \min(a, a) + \min(b, b) + 2(a, b) = \min(a, a) + \min(b, b)$

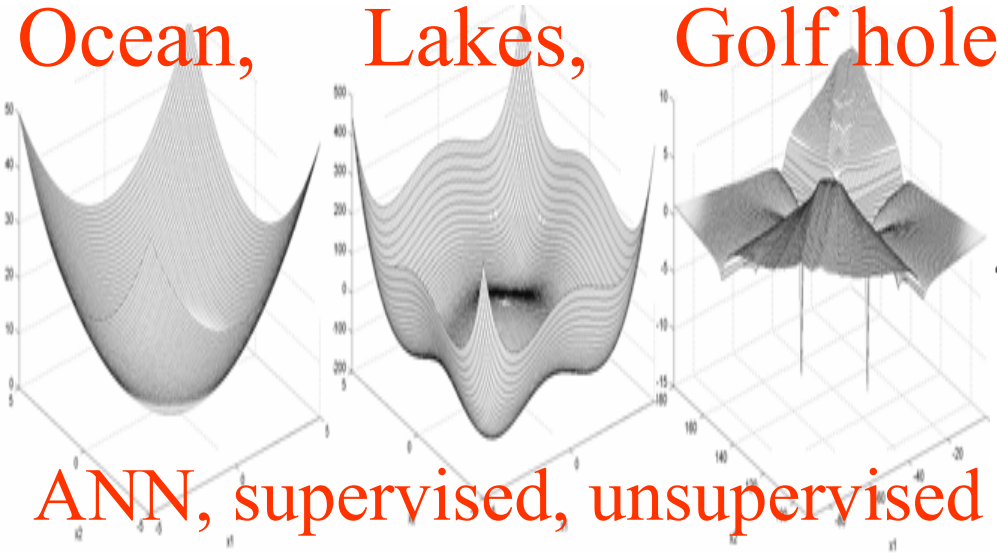
$$E = \frac{1}{2}\sum_i\sum_j\sum_k d_{i,j}(S_{i,k}S_{j,k+1} + S_{i,k}S_{j,k-1}) + c\sum_i(\sum_j S_{i,j} - 1) + c''\sum_j(\sum_i S_{i,j} - 1) + c'''\sum_i\sum_j S_{i,j}(S_{i,j} - 1)$$

Szu, International Joint Conf. on Neural Networks, Washington, DC, pp. I-507-511, June 1989.

Q.E.D.

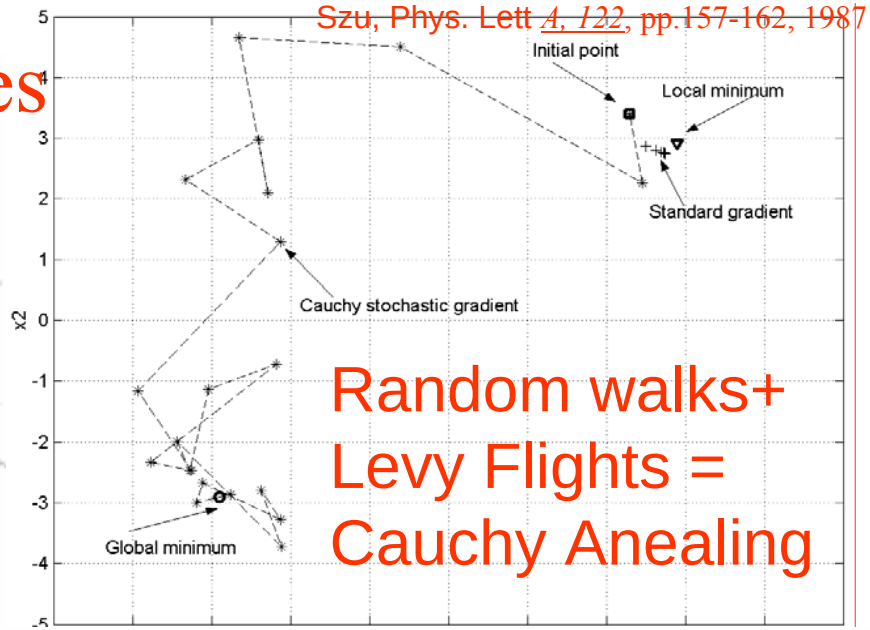
Learning Landscapes & Cauchy Annealing

Ocean, Lakes, Golf holes

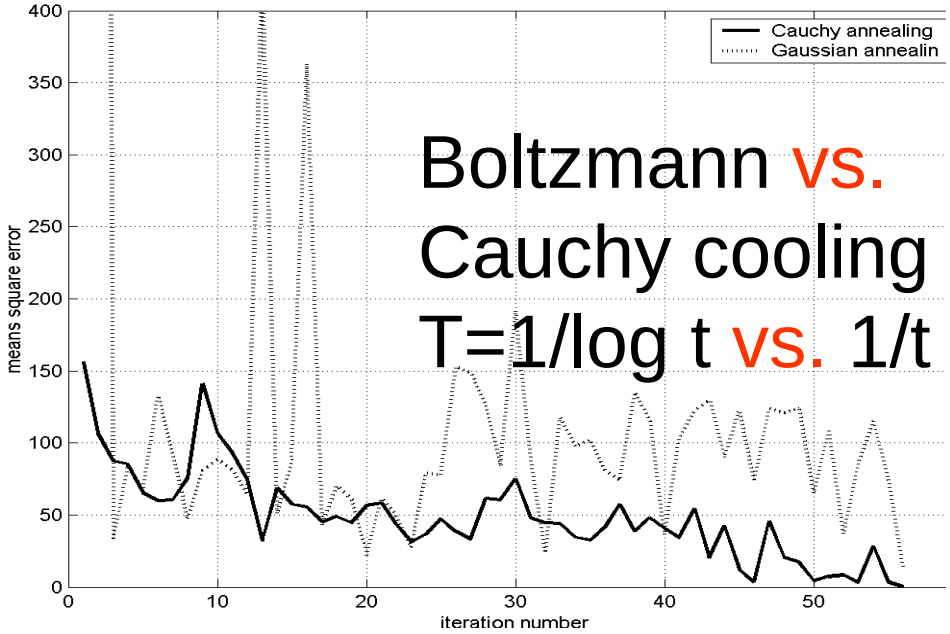


ANN, supervised, unsupervised

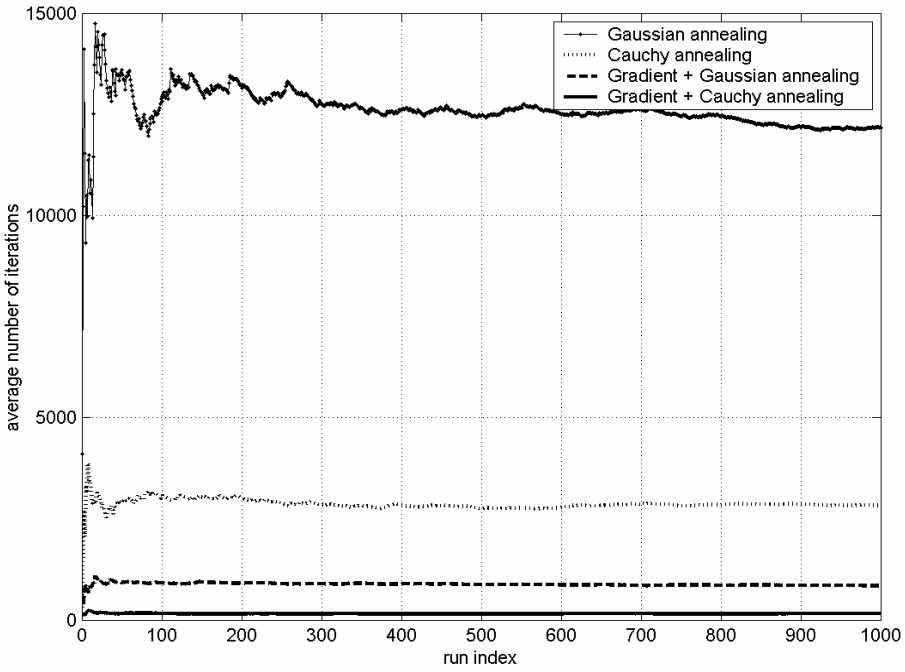
Szu, Phys. Lett *A, 122*, pp.157-162, 1987



Random walks+
Levy Flights =
Cauchy Anealing



Boltzmann vs.
Cauchy cooling
T=1/log t vs. 1/t



Theorem Cauchy vs. Gaussian cooling schedules

$$T_G(t) = T_0 / \log t \qquad T_C(t) = T_0 / t$$

Proof. By negate the converse strategy

Let the state-generating probability at the time t bounded below be g_t

Then the probability of not generating by $0 \le (1 - g_t)$

To prove a cooling schedule is to prove the impossibility of never generating a state vanished: equivalent to prove:

$$\prod_{t=t_0}^{\infty}(1-g_t) = 0 \quad (\log 0 = -\infty, \log(1-g_t) \approx -g_t) \quad \sum_{t-t_0}^{\infty} g_t = \infty$$

$$T_G(t) = T_0 / \log t$$

$$g_t \approx exp\left[-|\Delta x_0|^2 / T_G(t)\right] T_G(t)^{-D/2} \qquad \sum_{t=t_0}^{\infty} g_t \ge exp(-\log t) = \sum_{t=t_0}^{\infty} 1/t = \infty$$

$$T_C(t) = T_0 / t$$

$$g_t \approx \frac{T_C(t)}{\left[T_C^2(t) + |\Delta x_0|^2\right]^{(D+1)/2}} \approx \frac{T_0}{t|\Delta x_0|^{D+1}} \qquad \sum_{t=t_0}^{\infty} g_t \approx \frac{T_0}{|\Delta x_0|^{D+1}} \sum_{t=t_0}^{\infty} \frac{1}{t} = \infty$$

So neighborhood was visited infinite number of times at each time t for admissible cooling schedule.

Geman & Geman PAMI-6, pp. 721-741, Nov. 1984.
SZU & Hartley, Phys Lett. *A* 122, pp.157-162, 1987

Review of Stochastic Newtonian ANN Models & Convergence Proof

$$\frac{du_i}{dt} = F_i(t) \equiv \langle F_i(t) \rangle + \{F_i(t) - \langle F_i(t) \rangle\}; \quad \frac{du_i}{dt} + \eta u_i = \tilde{F}_i(t);$$

Langevin

$$\frac{du_i}{dt} + \eta u_i = -\frac{\partial E}{\partial v_i} + \tilde{F}(t); \quad v_i = \sigma(u_i); \quad u_i = \sum_j W_{ij} v_j - \theta_i \quad W_{ij} \propto v_i u_j$$

$$\frac{du'_i}{dt} = -\frac{\partial E'}{\partial v'_i} + \tilde{F}'(t); \quad v'_i = \sigma(u'_i); \quad u'_i = \sum_j W'_{ij} v'_j - \theta'_i; \; W'_{ij} \propto v'_i u'_j$$

$$u_i' \equiv u_i \exp(\eta t); \; v_i' \equiv v_i \exp(\eta t); \; E' \equiv \exp(2\eta t) E; \; \tilde{F}' = \exp(\eta t) \tilde{F}:$$

Input Hopfield dynamics

$$u_i - \sum_j W_{ij} v_j = 0; \Rightarrow {du_i}/{dt} = -\alpha(u_i - \sum_j W_{ij} v_j); \quad \Rightarrow {du_i}/{dt} + \alpha u_i = -\frac{\partial E}{\partial v_i}$$

Output Grossberg dynamics

$$v_i - \sigma(u_i) = 0; \quad {dv_i}/{dt} = -\beta(v_i - \sigma(u_i))$$

Hebbian ICA dynamics a la Amari-Cichocki-Yang metric

$$\frac{dW_{ij}}{dt} = -\frac{\partial \text{supervised } E}{\partial W_{\alpha\beta}} g_{ij}^{\alpha\beta}; \frac{dW_{ij}}{dt} = \frac{\partial \text{unsupervised } S/K}{\partial W_{\alpha\beta}} g_{ij}^{\alpha\beta} = 0 \Rightarrow f.p. \text{Fast ICA}$$

Convergence Theorem proved via Lyaponov style: (real#)2=positive

$$\frac{dE'}{dt} = \sum_j \frac{\partial E'}{\partial v'_j} \frac{dv'_j}{du'_j} \frac{\partial u'_j}{\partial t} + \sum_i \sum_j \frac{\partial E}{\partial W_{ij}} \frac{dW_{ij}}{dt}$$

$$= -\sum_j \left(\frac{\partial E'}{\partial v'_j} \right)^2 \frac{dv'_j}{du'_j} + \left(\frac{\partial E'}{\partial v'_j} \right) \frac{dv'_j}{du'_j} \tilde{F}'(t) - \sum_i \sum_j \left(\frac{\partial E}{\partial W_{ij}} \right) \frac{\partial E}{\partial W_{\alpha\beta}} g_{ij}^{\alpha\beta} \leq 0$$

Szu 1999 Q.E.D.

Comparison: Single pixel E([W]X-S) in isothermal equilibrium min. H=E-T_oS, versus ICA Max. Post Entropy for a fixed [A] for all pixel ensemble averaging

- *Space-variant imaging (Variable response [A] pixel to pixel)*
- *Pixel-parallel independent search*

Szu applied min. Helmholtz free energy per pixel to reduce the uncertainty among many inverse solutions.

Min. Helmholtz H=E([W]X-S) - T_oS(s)

Shannon-Boltzmann a priori Entropy:

$S(s_1,s_2,s_3,.)= -\Sigma s_i \log s_i+(\mu_o+1)(\Sigma s_i-1)$

by postulating the first order estimation error information I/O energy

$E=\mu([W]X-S)=\lambda([A]S-X)$

derived ANN sigmoid and Hebb rule:

Derived: 1. Sigmoid, 2. Unsupervised learning Hebb rule 3. Real world applications Remote sensing, breast cancer

- *Space invariant imaging (Identical response [A] of a closed system)*
- *advantage of pixel ensemble*
- *Bell-Sejnowski, Amari, Oja (BSAO) find ICA defined by joint pdf factorization*

$$\rho(x_1,x_2,x_3,.) \Rightarrow \rho_1(x'_1)\,\rho_2(x'_2)\ldots$$

$$\frac{\partial [W]}{\partial t} = \left\langle \frac{\partial\, PostEnt}{\partial [W]} (\vec{y}([W]\vec{x}))[W]^T[W] \right\rangle_{pixels}$$

$$d_{\vec{y}} = (\vec{y}, \vec{y}) = (\vec{x}, [W]^T[W]\vec{x}) = d_{\vec{x}}$$

BSAO assume ANN post-processing closed system

Max Post-Entropy(V = σ([W]X))

Challenges remain:

1. **Component Permutation?**
2. **Inhomogeneous pixel *[A]?***
3. **Nonlinear ICA?**
4. **Biological meaning of binding?**

References: Szu et al., SPIE Wavelets 1997; Szu et al.,WCCI Hawaii 2002; IEEE Trans Geoscience Remote Sensing 2002

To Be Continued…

SzuH@onr.navy.mil

703-696-6567 Harold Szu

Navy team of mini-UAVs

- Payload limitation of EO/IR sensors without Gimbals using HVS software de-jittering
- 3D Synthetic Aperture Radar at L band each mini-UAV uses GaAs cellular phone Transceiver as an aperture of EM(x,y,z,t)

Unique micro-UAV advantages are inexpensive, covert, distributive, and yet network-centric sensor suite fusion for surveillance com-link.

Its unique challenges are the size, weight, power, which translate to fixed sensors mount suffering the aerodynamics motions, jittering using low power on-board processing and C4ISR off-board decision aids.

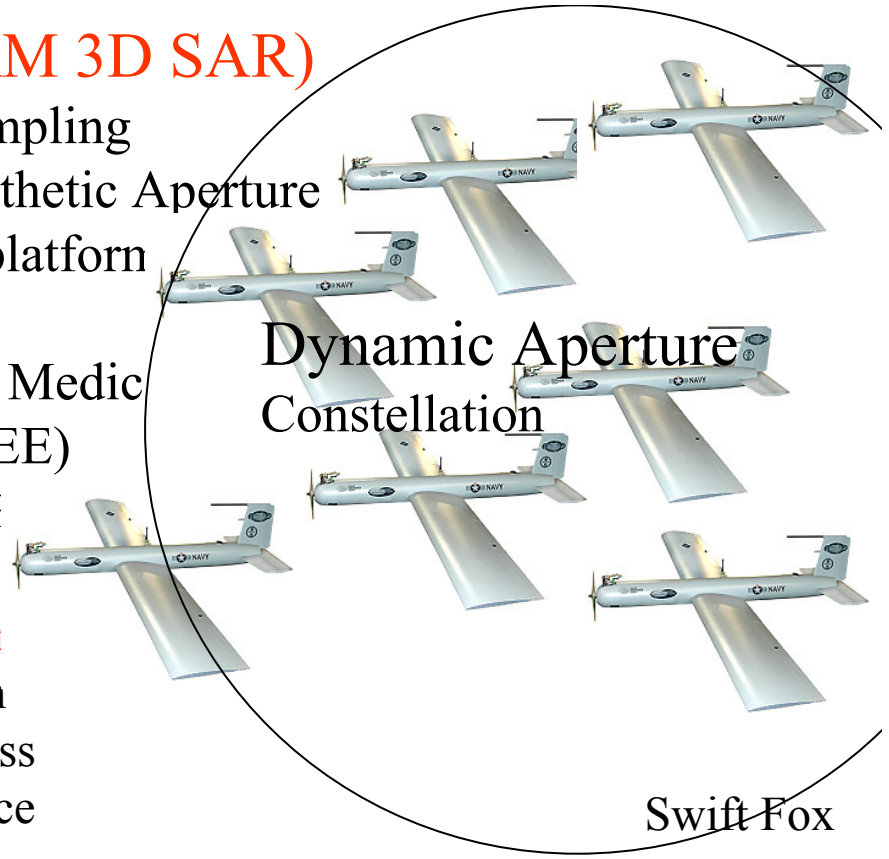
ONR Silver Fox (EO/IR & SWARM 3D SAR)

Enabling Law& Policies for sparse dynamic sampling of mobile transceiver elements as a team of Synthetic Aperture Radar (SAR) Speed of Light, >> small air platform

- FAA (low heights) <<1KM
- FCC (passive or low power)> wireless Medic (e.g. Blue Tooth, Home Appliance, IEEE)

Functionality of SWARM team of Small Silver F

- 3-D all Weather Surveillance & Imaging Foliage, Clouds, Fog, Smoke, Sand Stor
- Distributed Low-Cost SAR Constellation
- Close to Target Surveillance & Covertness
- Launch from Distance & Flight Endurance

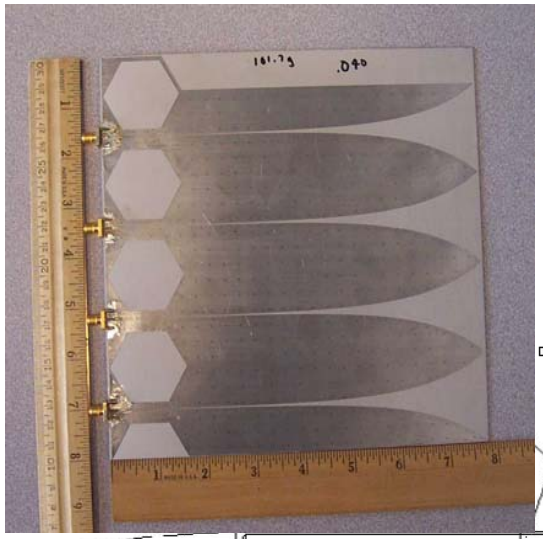


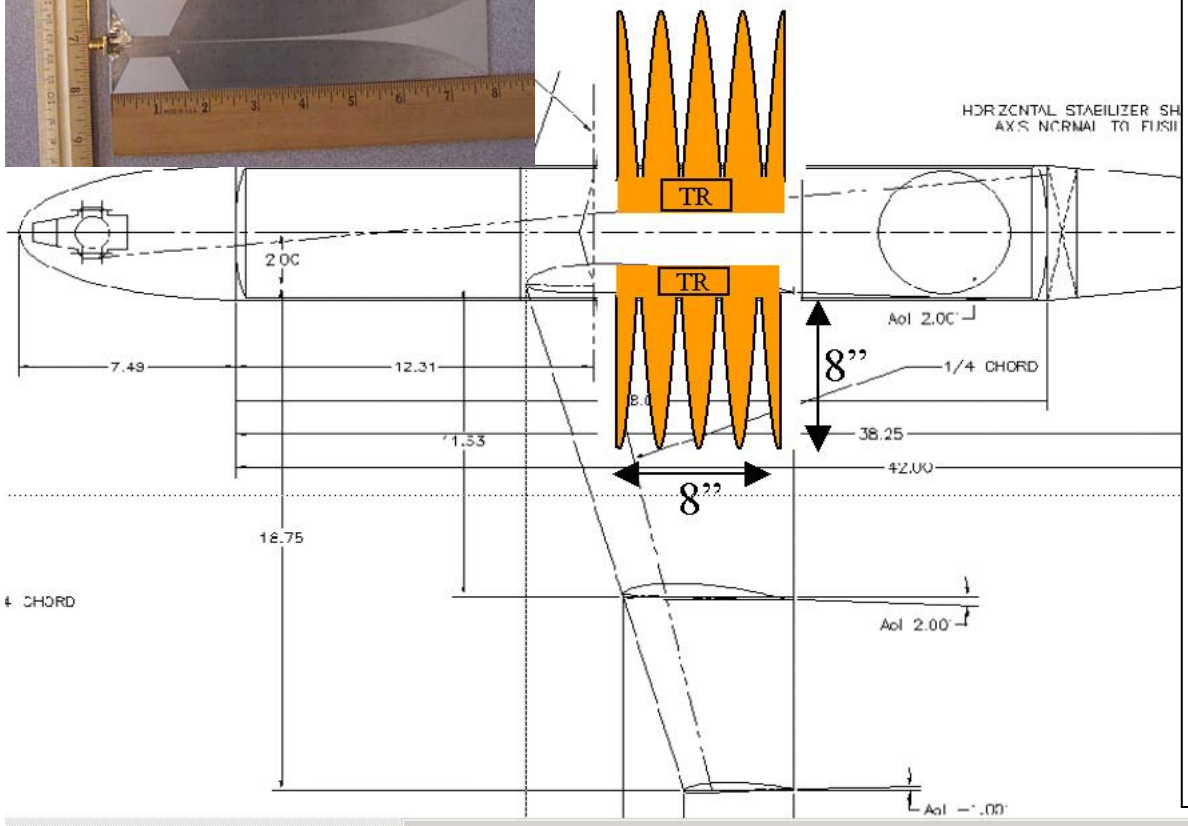
Enabling Technologies

- Cell Phone Technology ⇒ Ultra Low Cost Surface Mount L-Band RF Components

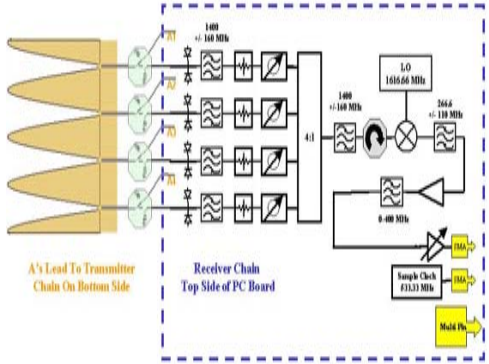
- SWARM UAV (Silver & Swift Fox) ⇒
  - Low Cost, Disposable (<$2K/bird)
  - Small (Size & Signature)
  - Long Duration Flight (up to 24 hours)
  - Payload Capacity (weight, power, size) for both EO and Radar Sensors

3D SAR WideBand Antenna & T/R

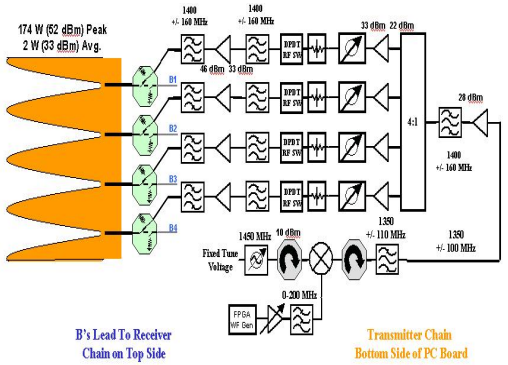




200 MHz BW Receiver Circuit

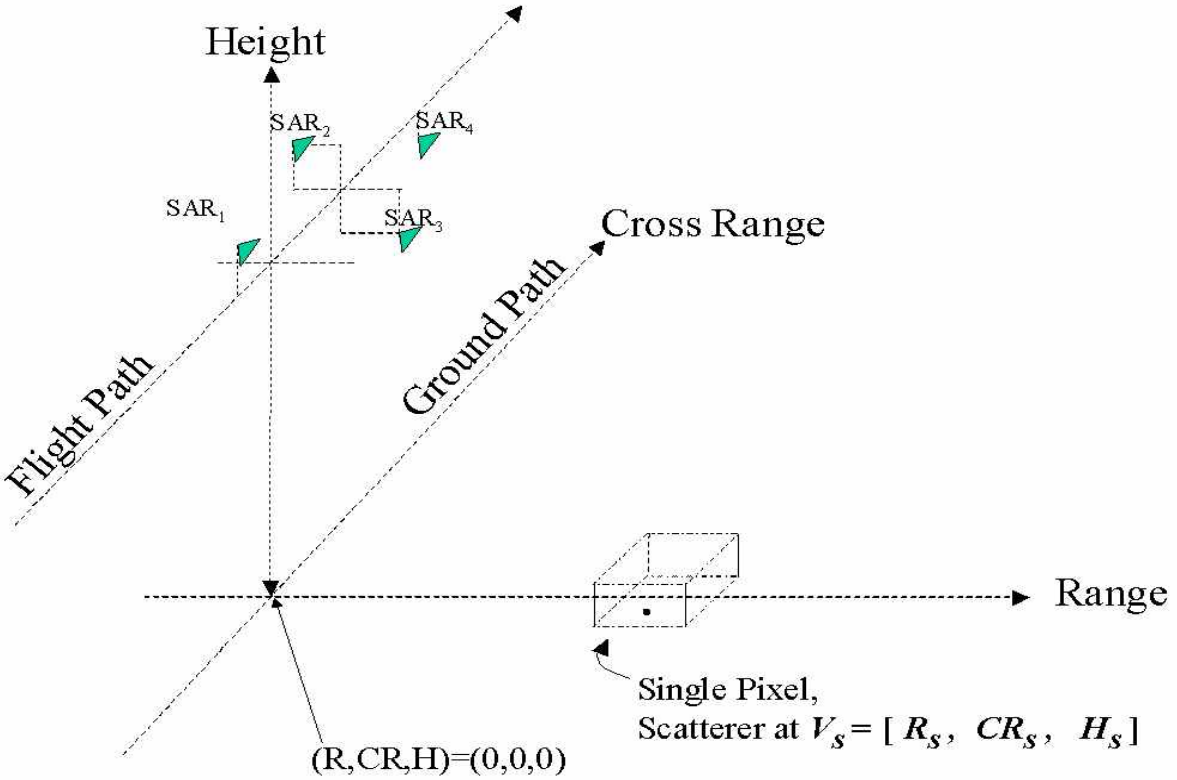


2W Avg Transmitter Circuit

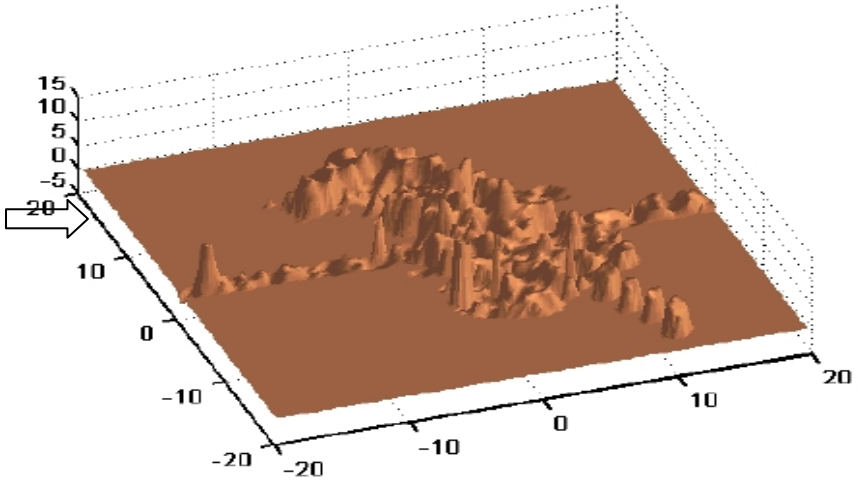


SWARM 3D SAR

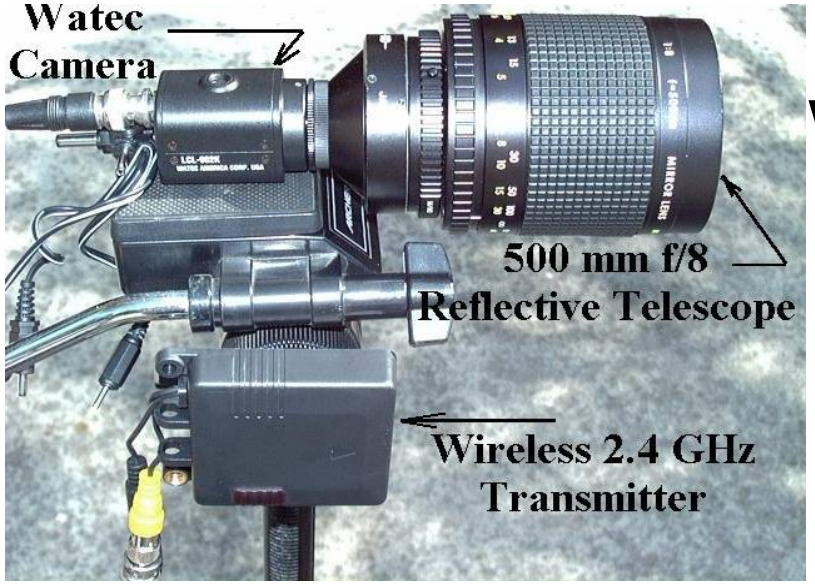
Low Cost Swift/Silver Fox
3-D SAR Sensor, Multi-platfor
0.8 meter resolution (2-D)
FOPEN (L-Band), 2 Watts avg
6 Km Swath Width



Goal (without the Zero Doppler Contamination):



Height Relief Map of a Tank Vehicle from Two Interferometric ISAR images (1 Foot Resolution)

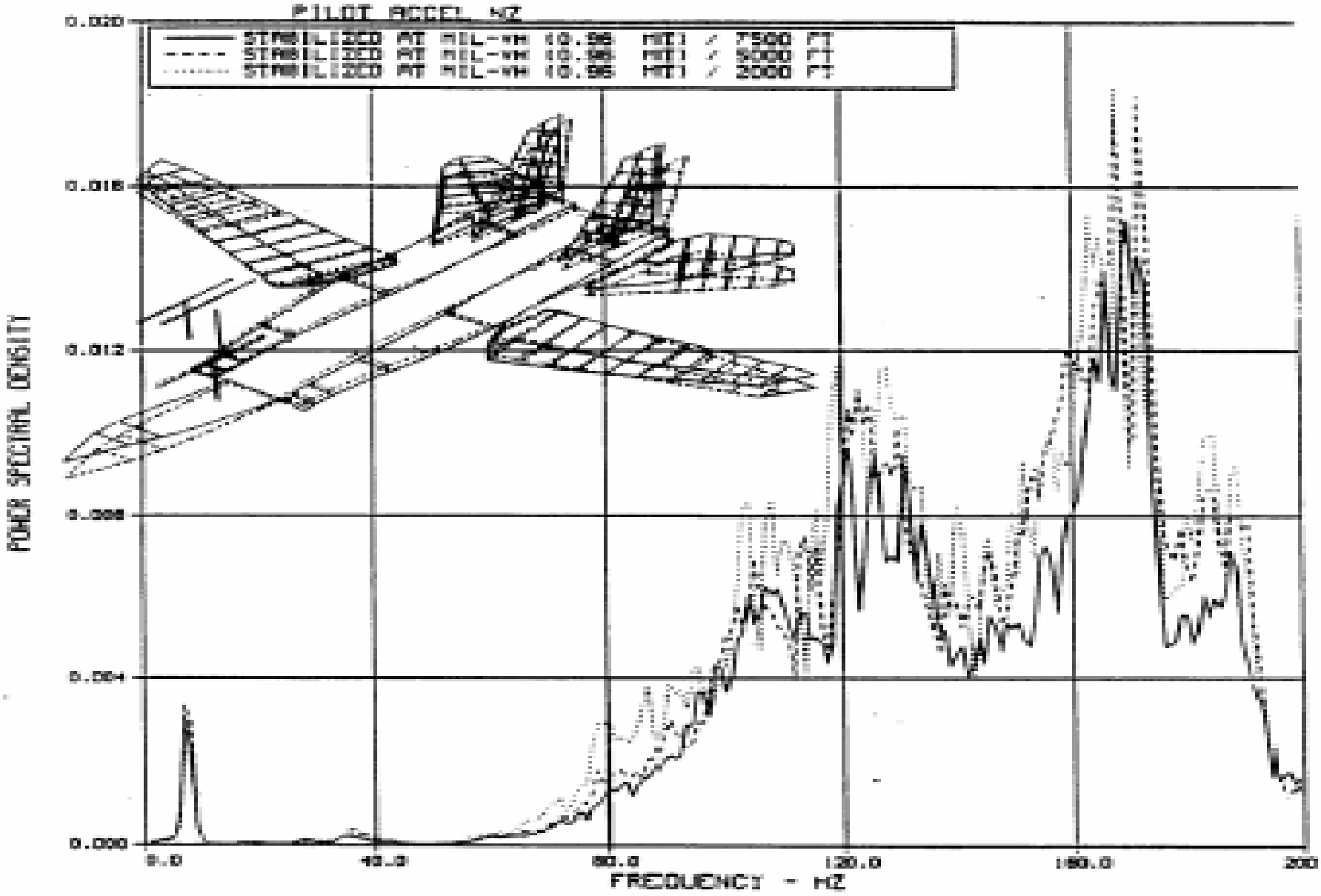


Why video sub-pixel jitter correction is important?

•We define jitter to be sub-pixel or small-amplitude vibrations up to one pixel, as opposed to motion blur over several pixels for which there already exists real time correction algorithms used on other platforms.

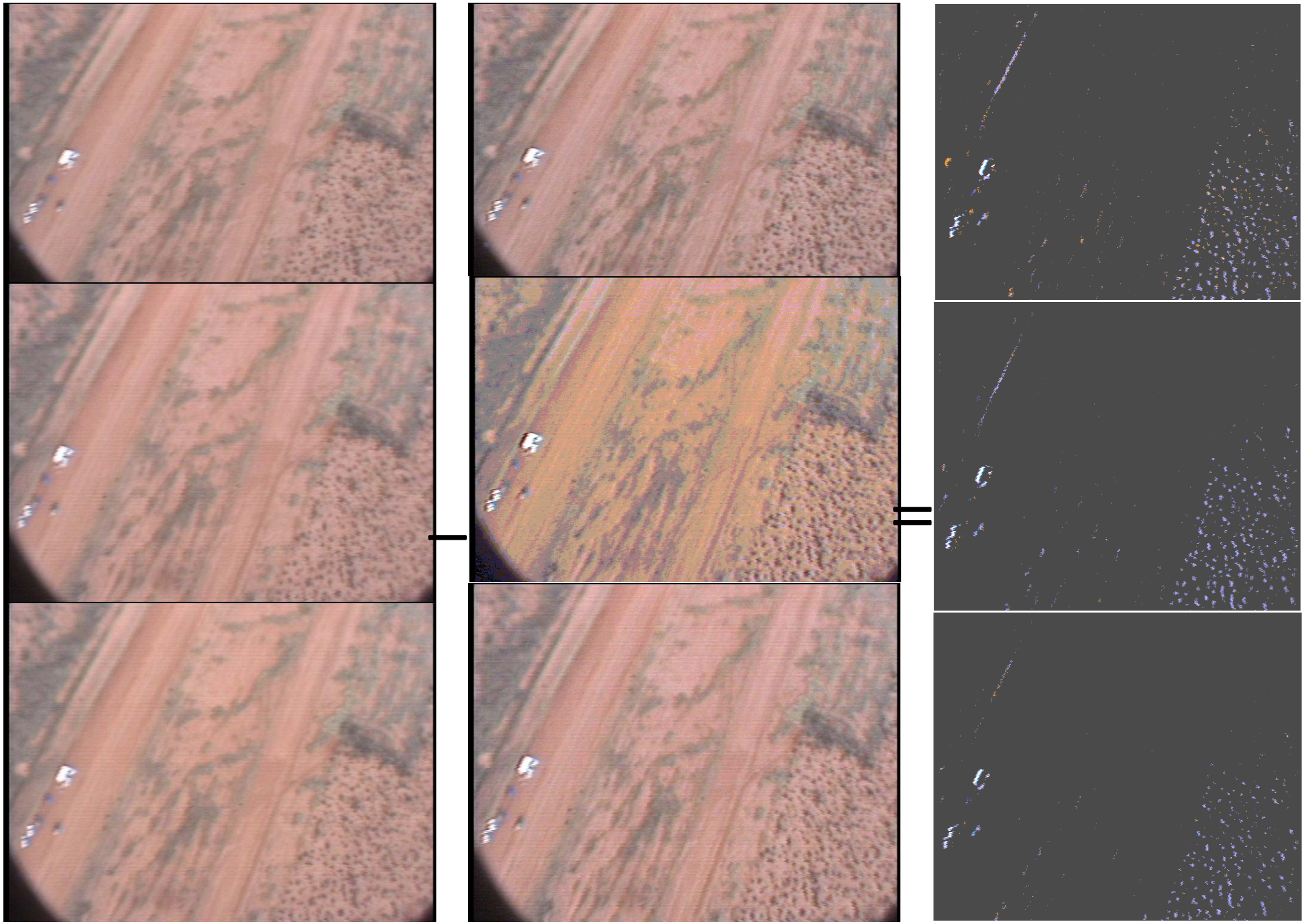
•Since micro-UAV, Silver Fox, cannot afford Gimbals mounting from the isolation coupling to the turbulent aerodynamics of the airframe, we must explore **real-time unsupervised learning software** on board of μ-UAV to mitigate the sub-pixel jitter effect.

•The sub-pixel accuracy is the basis of affine distortion transform and passive cell-phone transceiver array one per UAV for interference SAR registration.

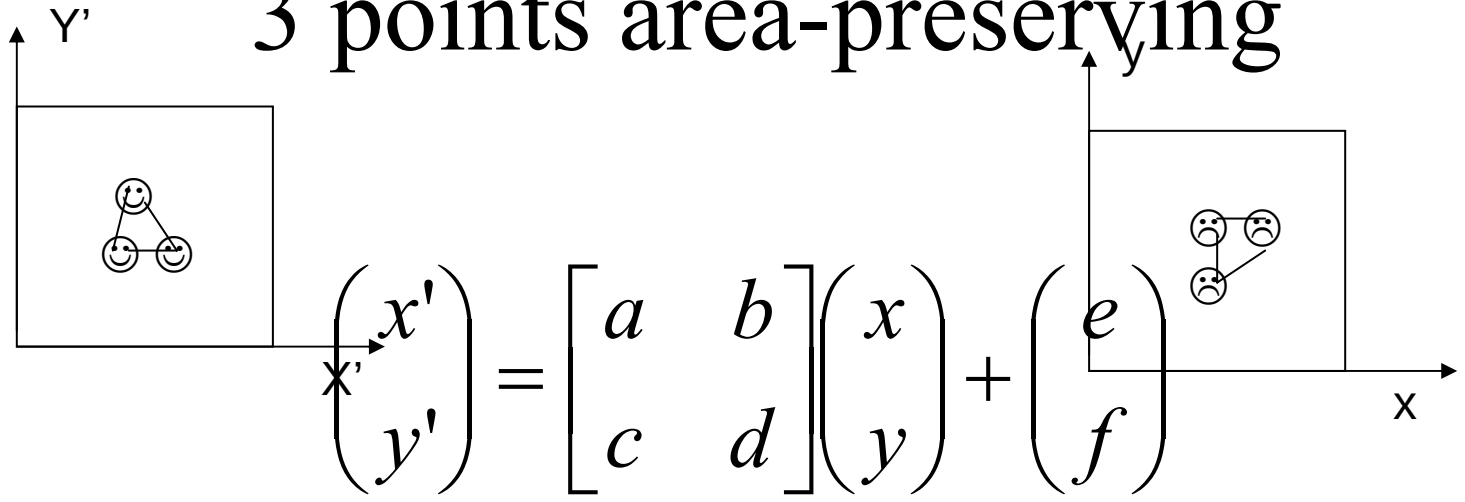


From pg 30 & 40 of
Grumman Aerospace Corp., Report No. LD-303D-89-002

μ-UAV sub-pixel jitter Algorithm:



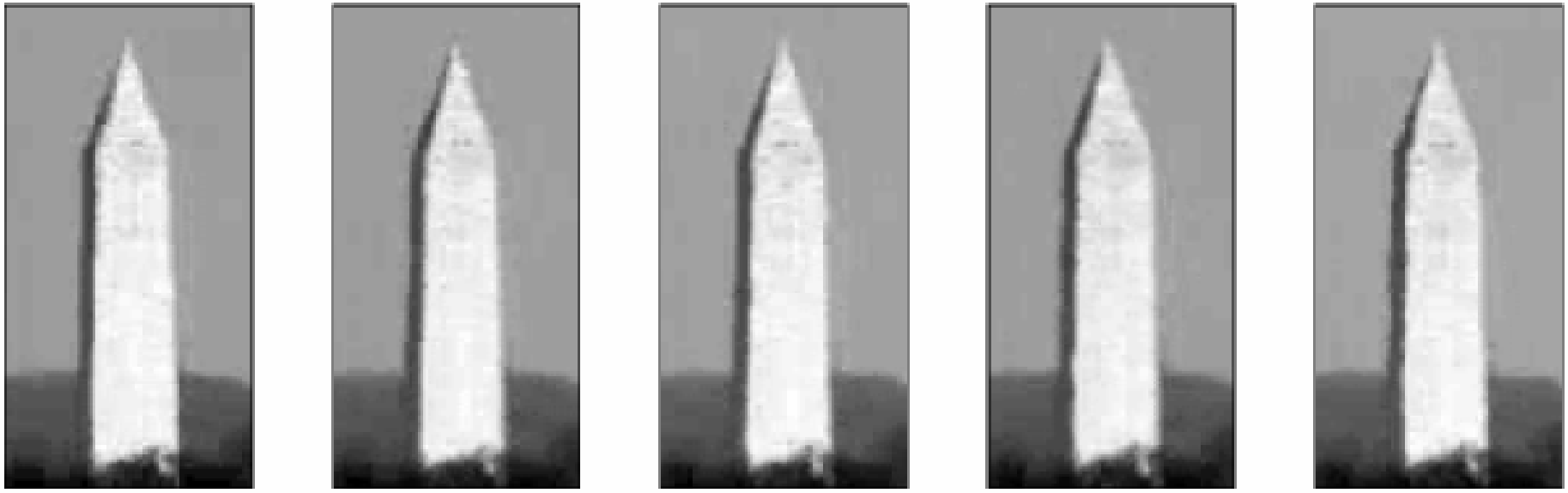
Cooperating Surveillance distortion registration by simple Affine Transform 3 points area-preserving



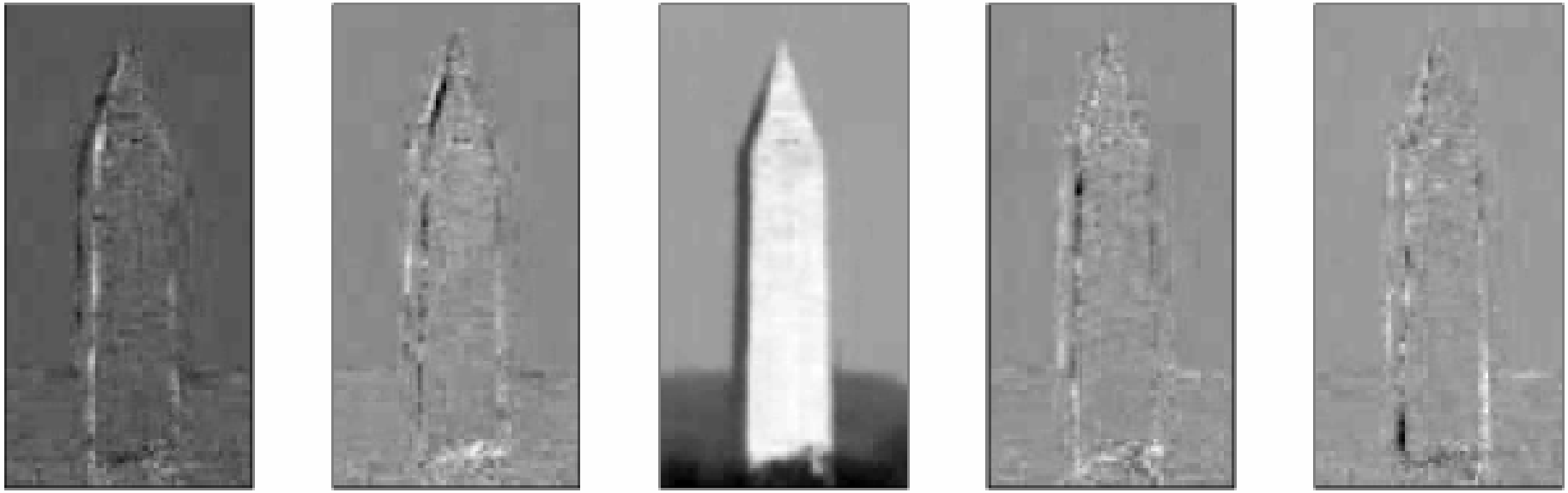
$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$$

Scatters of inexpensive plastic disks having both optical & thermal signatures provided the required minimum three correspondent points in neighborhood frames (over-determined case is given by Szu 1980). Then, location of approximate centers of scatters is geo-registered.

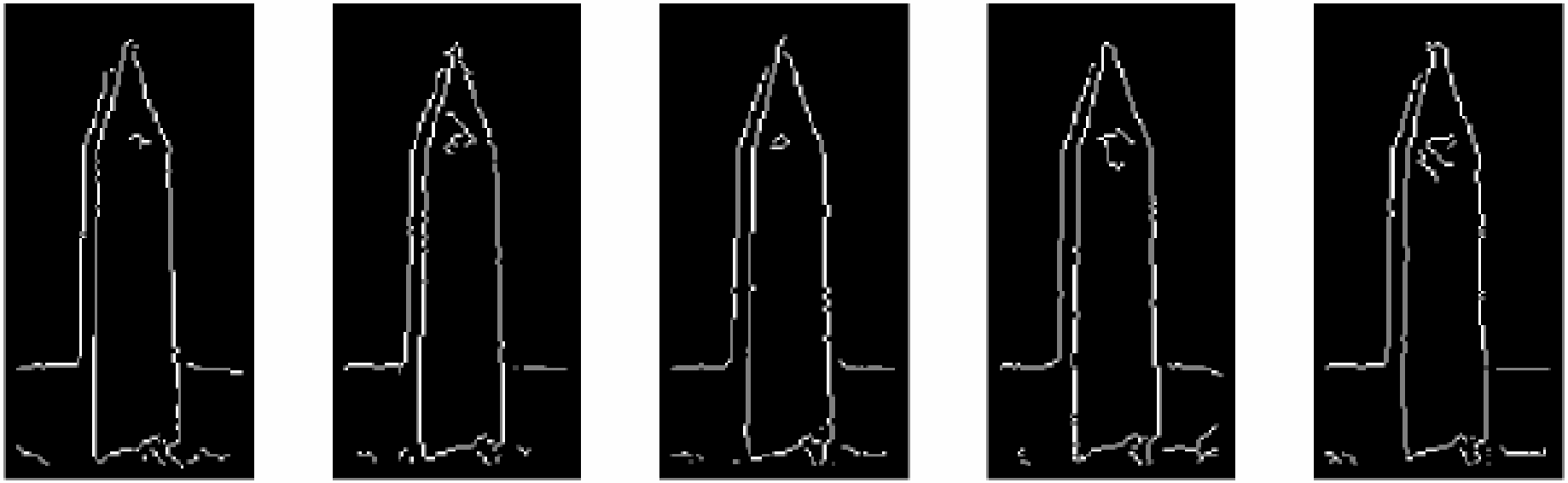
Data images



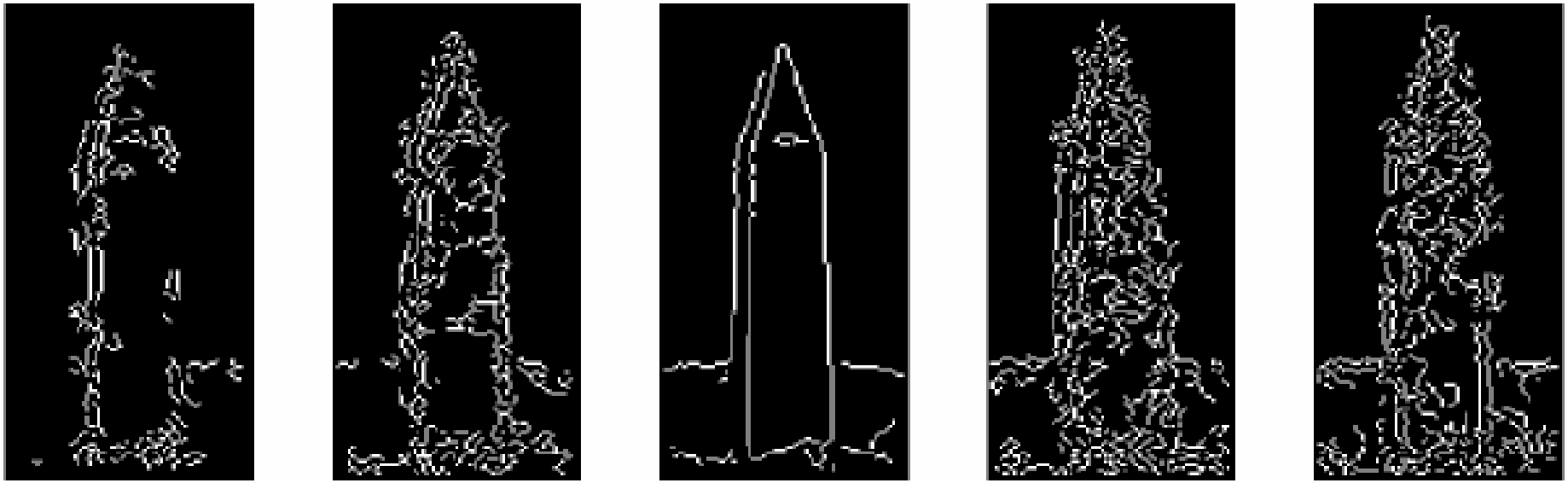
Source images



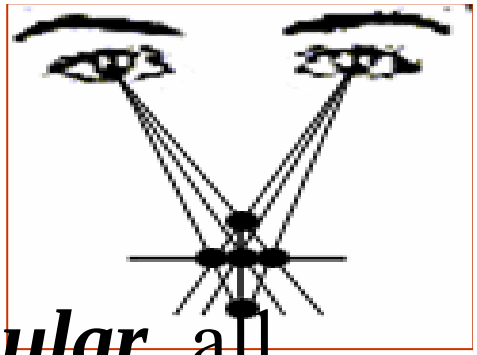
Data images - threshold=0.06



Source images - threshold=0.128



•Key ingredient of intelligent learning!—unsupervised learning via Power of Pairs.



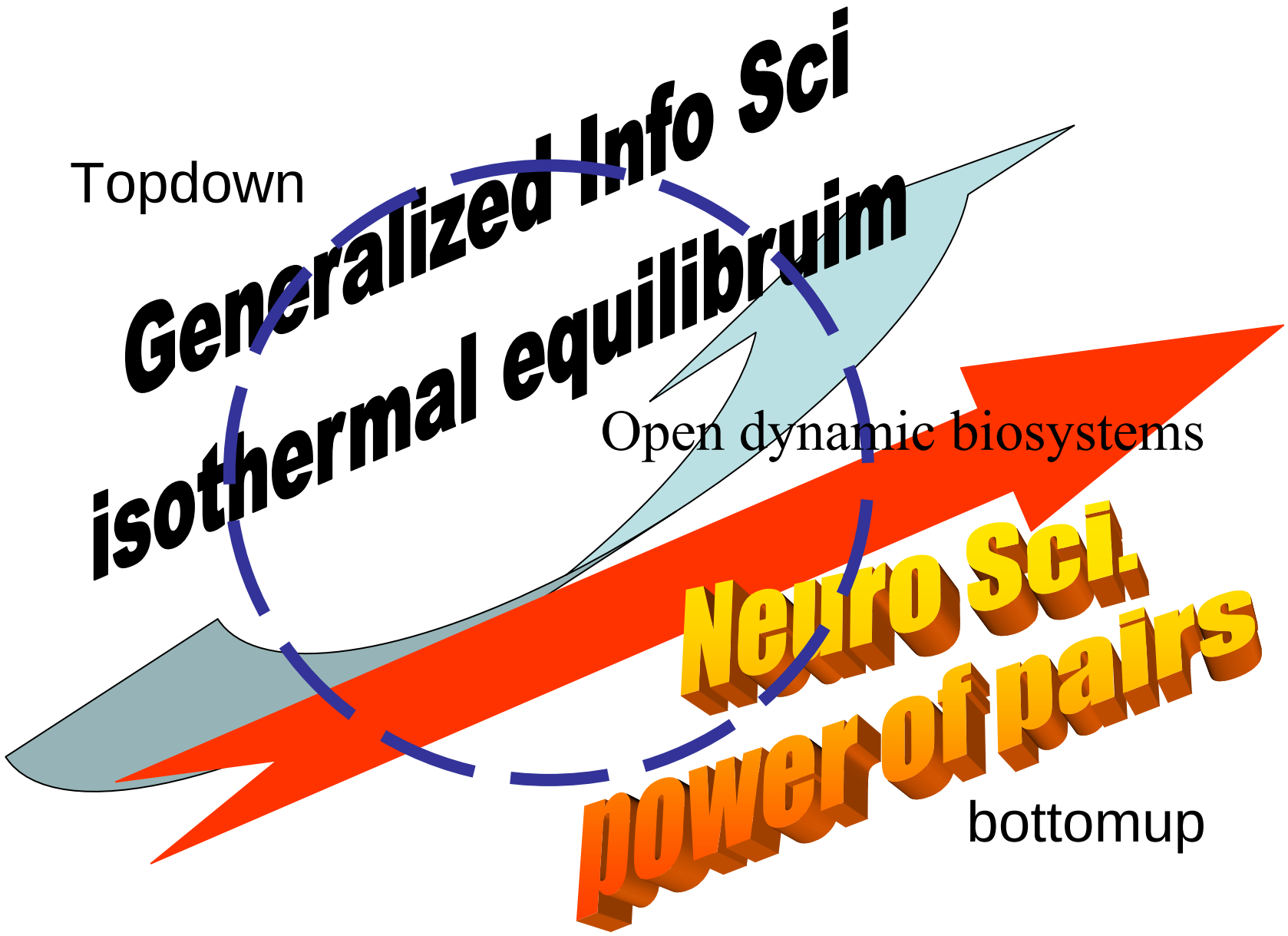
•While animal sensory ***outputs*** are ***singular***, all animal sensory *inputs* are from *pairs*!----two eyes, two ears, two nasals, tongue pair tessellations, two hands, etc.

•Answer:

Hardware fault tolerance and SNR enhancement by degeneracy coincidence account.

•Yet, current design of hearing aids remains to be indiscriminating signal from noisy and echoes without taking the advantage of power of pair.

Learning is a hallmark of natural intelligence (NI), understanding of unsupervised is a key breakthrough



We begin with two basic observations to generalize Shannon info theory to brain info theory

- ***What's evolutional advantages of keeping our brain at T_o=37°C?***
- cf. a warm blood *homosapiens* vs cold-blood dinosaur. Imagine long-term effect that internet computer is always-on to learn (e.g. Google) & to house keep (equivalent glial cells) all the time.

- ***Why do animals have pair inputs eyes, ears, etc. for single output?*** *Answer: Not simple as you think*. Unsupervised fusion

Neuroscience deals chaotic firing rates while Stat Mech deals vector time series. X(t)=[A(t)?]S(t)?

Why combine both ? Both address phenomenological ill-condition inverses. To invert matrix without knowing [A?] Given effects X(t)=[A?]S(t)? Find courses Physics demanding equilibrium: min. H=E –T_oS (isothermal free energy) can help select the unique S(t) for expt. X(t).

Riemannian Geometry Locally flat Stat Mech Physics

Poincare Map, an frozen order set generates recursive Flows of a vector time series of chaos

Fusion Novelty Detection

Blind **S**ources **S**eparation (not ICA)

US Patent Navy Case 8380 Szu, et al. on thermodynamics 4 equations.

Unknown source components $\{s_i\}$ inverting unknown matrix [A] by ANN weight matrix [W] have numerous solutions pairs sensor data vector $\mathbf{X = [A?]\ S?}$ *per pixel* (1)

The unique one is postulated to the minimum thermodynamic free energy $\boldsymbol{H}$ pixel-by-pixel min. $\boldsymbol{H = E - T_o S}$ (2)

The information energy is assumed by Szu as an analytical function of I/O in terms of the first order estimation error energy E valid for weak signals $E = \mu\,.([W]X - S?)$ (3)

(pixel-by-pixel μ-Lagrange vector): *minus the useless sources entropy of Shannon:*

$$italic\quad S = -\Sigma_i s_i \log s_i + (\mu_o + 1)(\Sigma_i s_i - 1) \quad (4)$$

(plume temperature parameter T_o is determined by a uniqueness requirement of the minimum H)

Software might also need a degree of M_IQ for de-conflicting incompatible programs or OS subsystem upgrade

Degree of difficulty:

1. Massive Parallel Processing
2. Applications domain integration
3. Self-diagnosis: Bug discovery program self-repair
4. Change cortex: Operation subsystems upgrades

All men are created equal;but machines are not.

Each device is designed differently to do specific functionality, a fuzzy membership function, e.g.
F ={JD(job description),M-M interface,
DC(duty cycle), Reliability, Alternative Power, integration with other device,…}
with a mean value that evolves in models or time to serve us.

Take F-ensemble Average Person Doing (APD):

$<APD>_F$ = Unity, what's half?

then M_IQ = a monotonic scale of $<APD>_F$

Could we agree M_IQ <50% supervised or rule-based?

Machines IQ Index

(1)MIQ ⊆ 50%. Supervised category with a lookup table having the extrapolation and interpolation capability up to MIQ ⊆ 50%.

(2) **MIQ beyond 50%.** Human-like sensor learning without supervision is believe to have half normal IQ, scored MIQ beyond 50%. Total combined Swarming Intelligence for self organization M_IQ is assumed to be unsupervised.

(Other than factory robots, futurist robots happen in an open, uncooperative and hazardous environment, e.g. deep ocean, outer space, melt down reactor) with an unforeseeable NL dynamics interwoven with non-stationary complexity).

MIQ=10% is loyal to its human master and its own survivability to differentiate electric power plug having a two-porn's of 110 Volts or three porn's of 220 Volts.2.

1. **MIQ=20%** is able to understanding human conversation in a fixed semantic network for a closed domain dialogue).
3. **MIQ=30%** is able to read facial expression and voice tone for e-IQ to understanding the emotion need of human being.
4. **MIQ=40%** is able to command and control a small team of other robots.
5. **MIQ=50%** is able to "explore the tolerance of imprecision," e.g. using fuzzy logic to negotiate a single precision path finding in an open save terrain.

Information Degree of Freedom

- Basic info-assurance question is how can one be sure about the trustworthiness of any info. Sirovich (PNAS 2003) analyzed the past 8 years of 9 US Supreme Court Judges, whose complete record of votes +1, -1, abstained 0 over 640 cases of which 448 are admitted.
- 9 judges in 9 D: [Breyer, Ginsburg, Kennedy, O'Connor, Rehnquist, Scalia, Souter, Stevens, Thomas] = [-1, -1, 1, 1, 1, 1, -1, -1, 1] for 2000 US President election
- PCA shows unprejudiced degree of freedom is 4.68 ideal judges.
- Two party political system supports 9 judges which might be only camps. In reality it seems to be 4 & half parties.

Smart Vision Sensors: Rods & Cons

- Massive parallel & densely uniformly packed fovea 6.3 Millions cons for color perception
- Distribution of 150 Millions rods are non-uniformly, polar-exponentially packed along radial direction
- Scale Invariant Architecture-Algorithm (Archirithm or Algotecture): data X = exp (U), feature domain u' =log x/2= log x- log 2=U gracefully degradation
- Ultra sensitivity: How could Single Photon be Detected by Human Vision Sensor at dark night? Although Rhodopsin molecule convert efficiently photons to charges as signal, but the energy of a photon is not enough for detection.

Single Photon Detection Design Logic

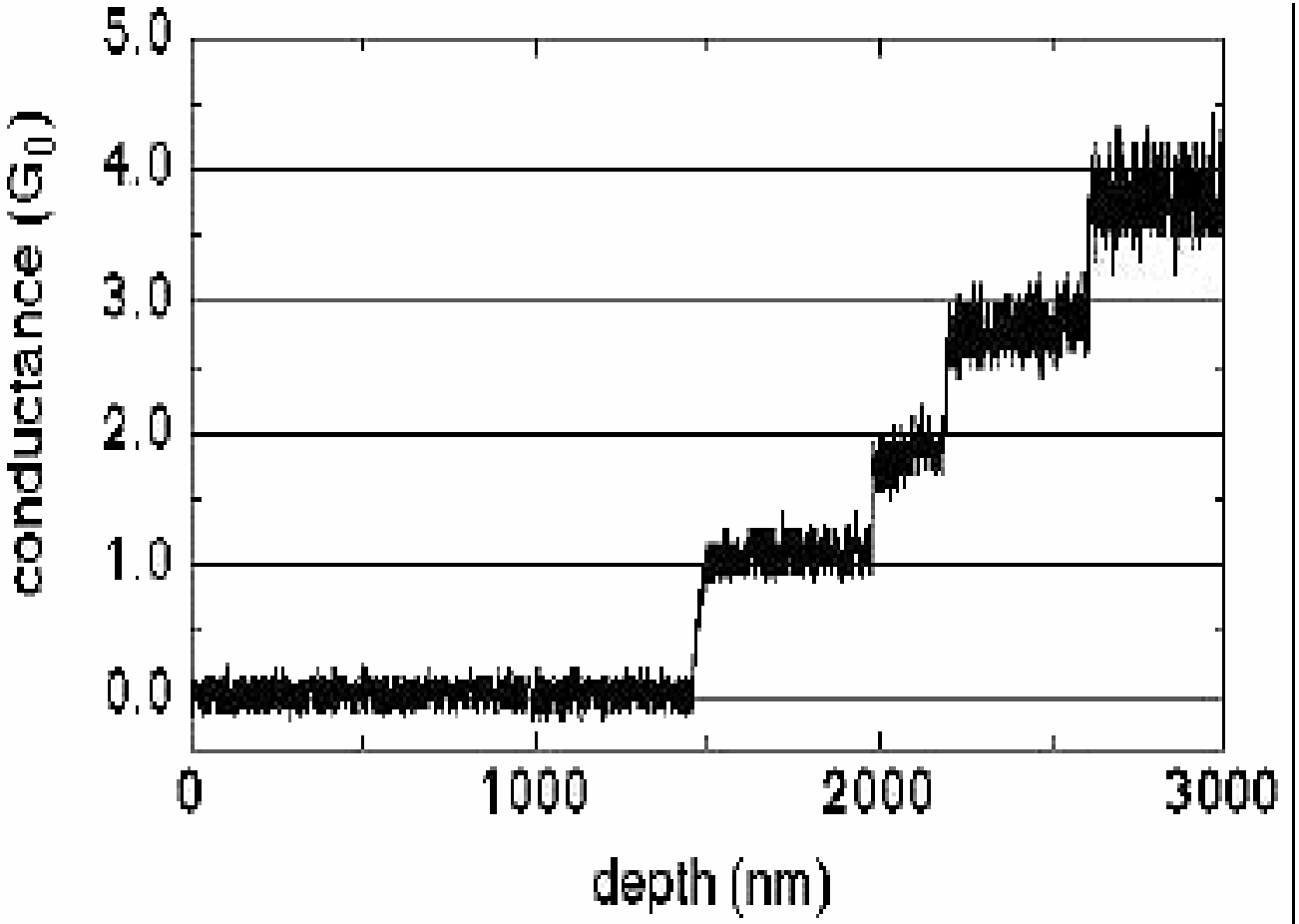
- Design Logic by means of “negate the converse”, “-1x-1=+1”i.e. a single photon can not provide enough signal charge over noise (SNR) but it can nevertheless affect the photoelectric medium to tip or switch the balance of an already existed & balanced so-called dark current driven by internal energy toward a different path, e.g. Wheatstone bridge of 4 arms in balance.

Middle Infrared Implementation Material

- Biomimetic Implementation: 1-D Quantum Carbon Nanotubes (CNT) in zigzag crystalline state enjoys bandgap semiconductor, which can be tuned by radius selectively at MidIR spectral.
- Working Hypothesis: CNT should suffer less thermal noise owing to a restricted geometry of noisy phonon excitation (TBD).

1-D multi-walls CNT Electrical Transport in Armchair

- Ballistic quantum conductance
- Absence of backward scattering
- High maximum current density: 10^7~10^13 A/cm^2

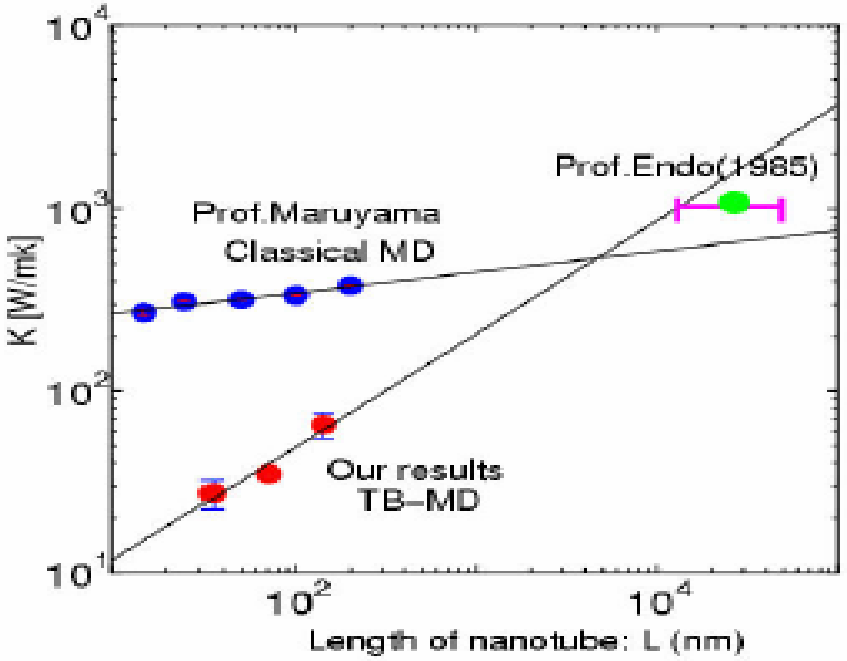


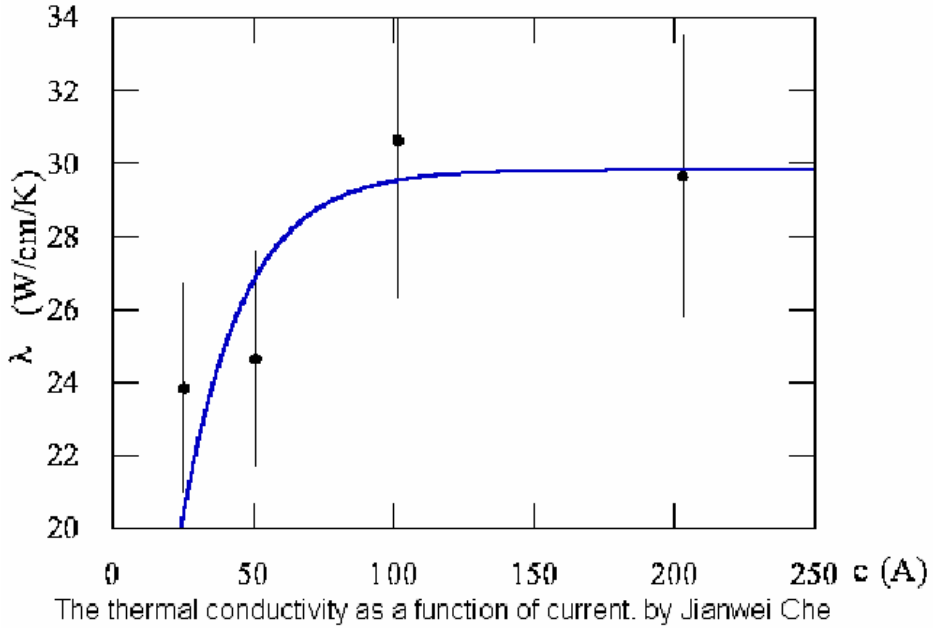
Above: A carbon nanotube is a quantum conductor. The conductance rises by 1 G_0 as the depth increases sufficently. G0= 2e2/h

Stefan Frank et al., Science 280 1744 (1998)

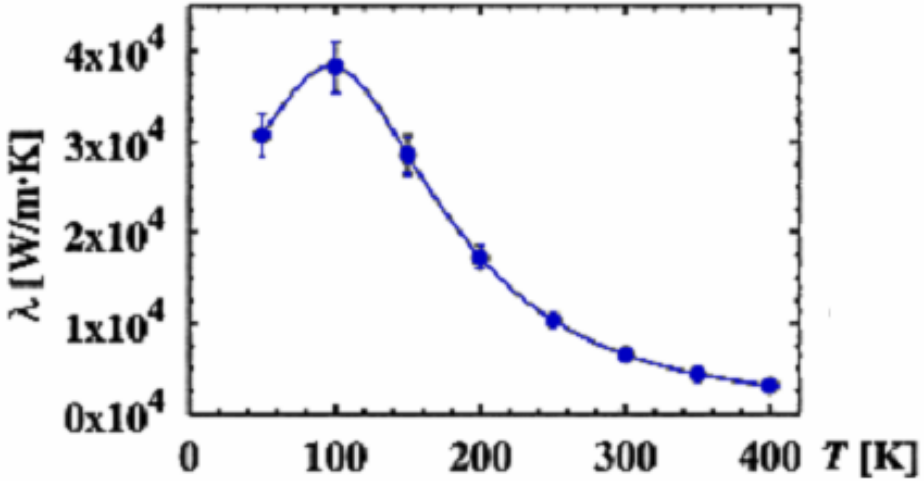
Thermal Transport in CNT

- Temperature dependency
- Length dependency
- Long-wave phonon: twiston





The thermal conductivity as a function of current. by Jianwei Che



Thermal Conductivity as a function of Temperature
(S. Berber, et al 84, 4613(2000))

Turbulence References

1. L.C. Andrews, R.L. Phillips, and C.Y. Hopen, *Laser Beam Scintillation with Applications*, SPIE Engineering Press, 2001.
2. L.C. Andrews and R.L. Phillips, *Laser Beam Propagation through Random Media*, SPIE Engineering Press, 1998.
3. L.C. Andrews, R.L. Phillips, C.Y. Hopen, and M.A. Al-Habash, *J. Opt. Soc Am. A*, 16, 1417-1429, 1999.
4. A. Consortini, F. Cochetti, J.H. Churnside, and R.J. Hill, *J. Opt. Soc. Am A* 10, 2354-2362, 1993.
5. L. C. Andrews, *Special Functions of Mathematics for Engineers*, Oxford University Press, 1998.
6. J.W. Strohbehn and R. S. Lawrence, "A Survey of Clear-Air Propagation Effects Relevant to Optical Communications," *Proc. of IEEE*, **58**, 22-44, 1970.
7. V.P. Lukin and V.V. Pokasoc, *Appl. Opt.,* 20, 121-135, 1981.
8. G.M.B. Bouricius and S.F. Clifford, *J. Opt. Soc Am*, 60, 1484-1489, 1970.
9. A. S. Gurvich and M.A. Kallistratova, "Experimental Study of the Fluctuations in Angle of Incidence of a Light Beam Under Conditions of Strong Intensity Fluctuations,"*Radiofizika*, 11, 66-71, 1968.
10. C. Y. Young, A. J. **Masino**, and F. Thomas, "Phase Fluctuations in Moderate to Strong Turbulence," Proc. of SPIE, 4976, 141 – 148 (2003).
11. C. Y. Young, A. J. **Masino**, F. E. Thomas, and C. J. Subich, *Atmosphere Induced Frequency Fluctuations*, submitted, Optical Engineering.
12. C. Y. Young, A. J. **Masino**, F. E. Thomas, and C. J. Subich, "The Wave Structure Function in Weak to Strong Fluctuations", WIRM, **14**, 75-96 (2004).
13. A. J. **Masino**, C. Y. Young, and F. Thomas, "Moderate-Strong Theory Applied to Second Order Statistics," Proc. of SPIE, 4976, 188 – 194 (2003).