Image processing and analysis of digital data and cryo-EM

Leandro F. Estrozi

Oleron - june 2019

What is a byte?
How numbers are stored in a computer?
What is a pixel?
How pixels are stored as an image?
What is an image? (for a SCIENTIST)
Digital signal processing with Mr.Fourier
1 dimension, 2 dimensions, 3 dimensions
Symmetry!

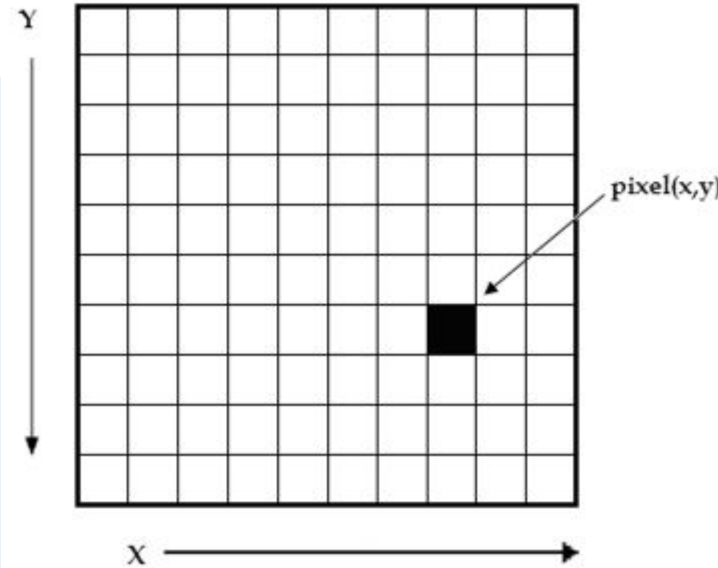
Image sampling and quantization

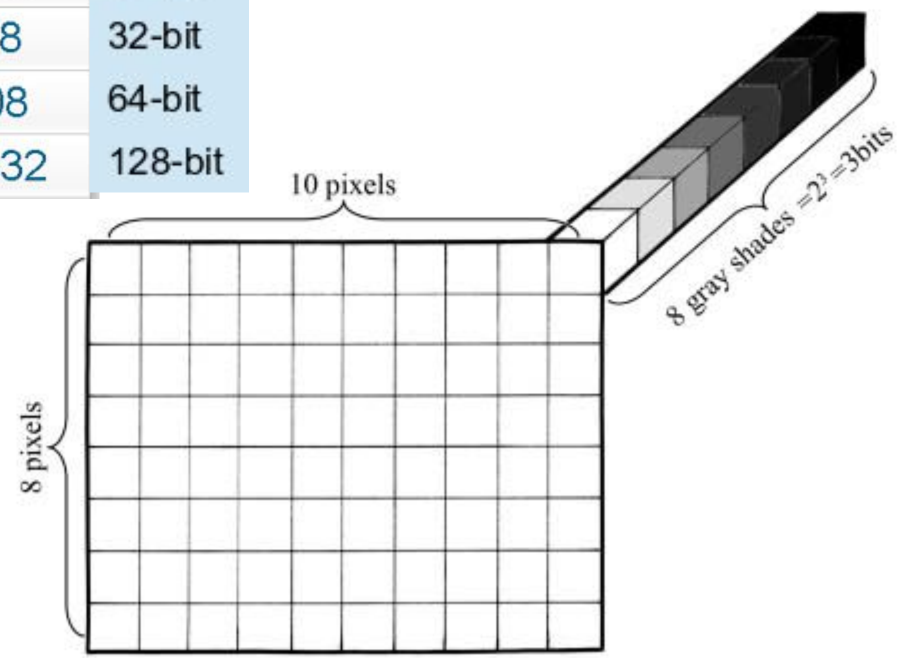
DATA TYPE	MIN_VALUE	MAX_VALUE	
unsigned char	0	255	8-bit
signed char	-128	127	8-bit
unsigned short int	0	65535	16-bit
signed short int	-32768	32767	16-bit
unsigned int	0	65535	32-bit
signed int	-32768	32767	32-bit
unsigned long int	0	4294967295	64-bit
signed long int	-2147483648	2147483647	64-bit
float	-3.4 *10^-38	3.4 *10^38	32-bit
double	-1.7*10^-308	1.7*10^308	64-bit
long double	-3.4*10^-4932	1.1*10^+4932	128-bit

From:
https://www.ibm.com/support/knowledgecenter/en/SSLTBW_2.1.0/com.ibm.zos.v2r1.cbcpx01/datatypesize64.htm

Image sampling and quantization

DATA TYPE	MIN_VALUE	MAX_VALUE	
unsigned char	0	255	8-bit
signed char	-128	127	8-bit
unsigned short int	0	65535	16-bit
signed short int	-32768	32767	16-bit
unsigned int	0	65535	32-bit
signed int	-32768	32767	32-bit
unsigned long int	0	4294967295	64-bit
signed long int	-2147483648	2147483647	64-bit
float	-3.4 *10^-38	3.4 *10^38	32-bit
double	-1.7*10^-308	1.7*10^308	64-bit
long double	-3.4*10^-4932	1.1*10^+4932	128-bit





From: https://www.ibm.com/support/knowledgecenter/en/SSLTBW_2.1.0/com.ibm.zos.v2r1.cbcpx01/datatypesize64.htm

American Standard Code for Information Interchange (1963)

Dec	Hex	Name	Char	Ctrl-char
0	0	Null	NUL	CTRL-@
1	1	Start of heading	SOH	CTRL-A
2	2	Start of text	STX	CTRL-B
3	3	End of text	ETX	CTRL-C
4	4	End of xmit	EOT	CTRL-D
5	5	Enquiry	ENQ	CTRL-E
6	6	Acknowledge	ACK	CTRL-F
7	7	Bell	BEL	CTRL-G
8	8	Backspace	BS	CTRL-H
9	9	Horizontal tab	HT	CTRL-I
10	0A	Line feed	LF	CTRL-J
11	0B	Vertical tab	VT	CTRL-K
12	0C	Form feed	FF	CTRL-L
13	0D	Carriage feed	CR	CTRL-M
14	0E	Shift out	SO	CTRL-N
15	0F	Shift in	SI	CTRL-O
16	10	Data line escape	DLE	CTRL-P
17	11	Device control 1	DC1	CTRL-Q
18	12	Device control 2	DC2	CTRL-R
19	13	Device control 3	DC3	CTRL-S
20	14	Device control 4	DC4	CTRL-T
21	15	Neg acknowledge	NAK	CTRL-U
22	16	Synchronous idle	SYN	CTRL-V
23	17	End of xmit block	ETB	CTRL-W
24	18	Cancel	CAN	CTRL-X
25	19	End of medium	EM	CTRL-Y
26	1A	Substitute	SUB	CTRL-Z
27	1B	Escape	ESC	CTRL-[
28	1C	File separator	FS	CTRL-\
29	1D	Group separator	GS	CTRL-]
30	1E	Record separator	RS	CTRL-^
31	1F	Unit separator	US	CTRL-_

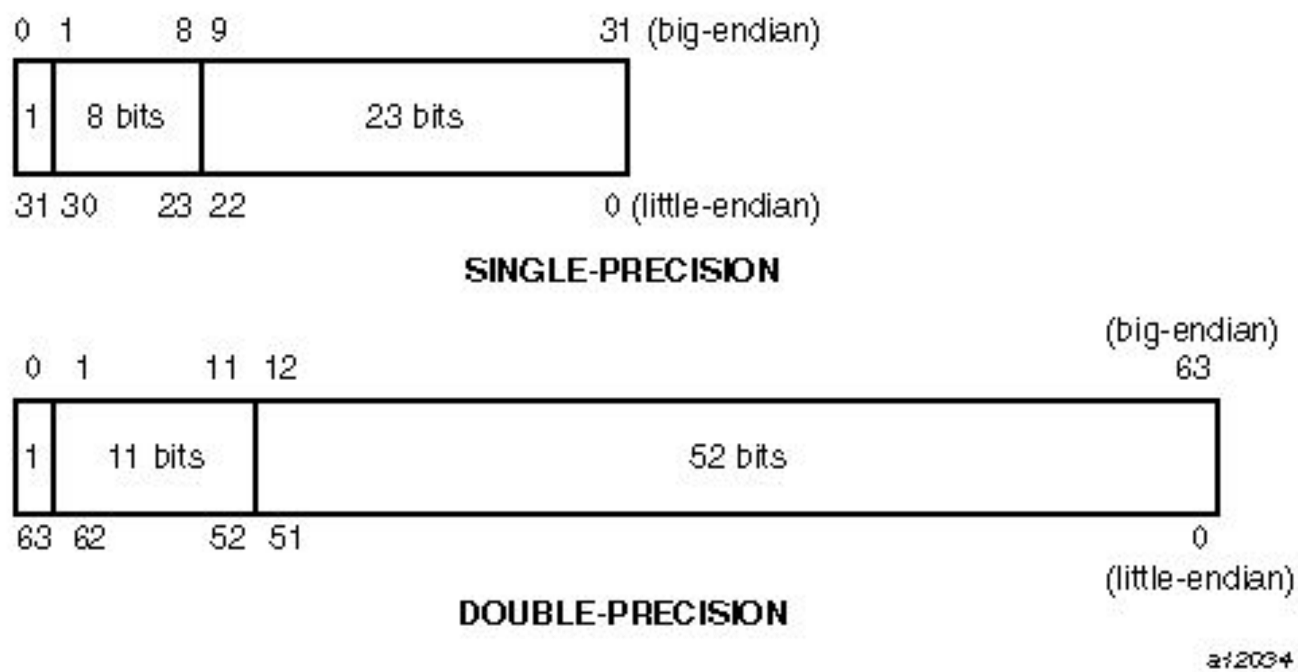
Dec	Hex	Char
32	20	Space
33	21	!
34	22	"
35	23	#
36	24	$
37	25	%
38	26	&
39	27	'
40	28	(
41	29	)
42	2A	*
43	2B	+
44	2C	,
45	2D	-
46	2E	.
47	2F	/
48	30	0
49	31	1
50	32	2
51	33	3
52	34	4
53	35	5
54	36	6
55	37	7
56	38	8
57	39	9
58	3A	:
59	3B	;
60	3C	<
61	3D	=
62	3E	>
63	3F	?

Dec	Hex	Char
64	40	@
65	41	A
66	42	B
67	43	C
68	44	D
69	45	E
70	46	F
71	47	G
72	48	H
73	49	I
74	4A	J
75	4B	K
76	4C	L
77	4D	M
78	4E	N
79	4F	O
80	50	P
81	51	Q
82	52	R
83	53	S
84	54	T
85	55	U
86	56	V
87	57	W
88	58	X
89	59	Y
90	5A	Z
91	5B	[
92	5C	\
93	5D	]
94	5E	^
95	5F	_

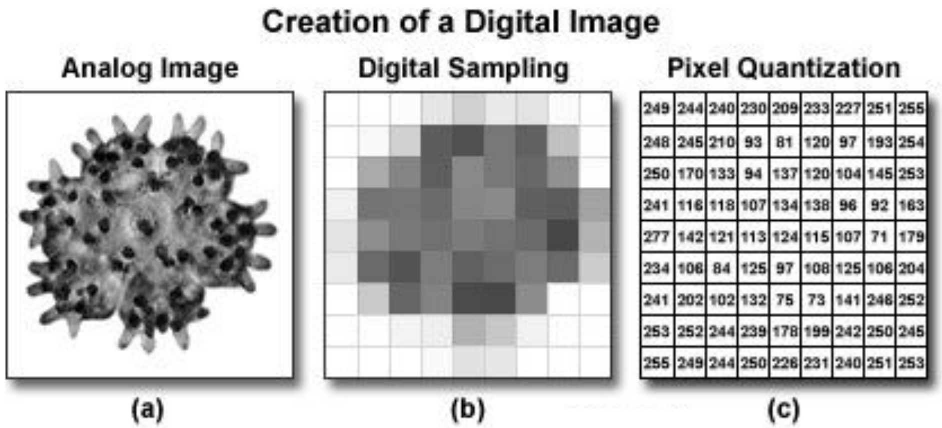
Dec	Hex	Char
96	60	`
97	61	a
98	62	b
99	63	c
100	64	d
101	65	e
102	66	f
103	67	g
104	68	h
105	69	i
106	6A	j
107	6B	k
108	6C	l
109	6D	m
110	6E	n
111	6F	o
112	70	p
113	71	q
114	72	r
115	73	s
116	74	t
117	75	u
118	76	v
119	77	w
120	78	x
121	79	y
122	7A	z
123	7B	{
124	7C	|
125	7D	}
126	7E	~
127	7F	DEL

Floating-point number representation

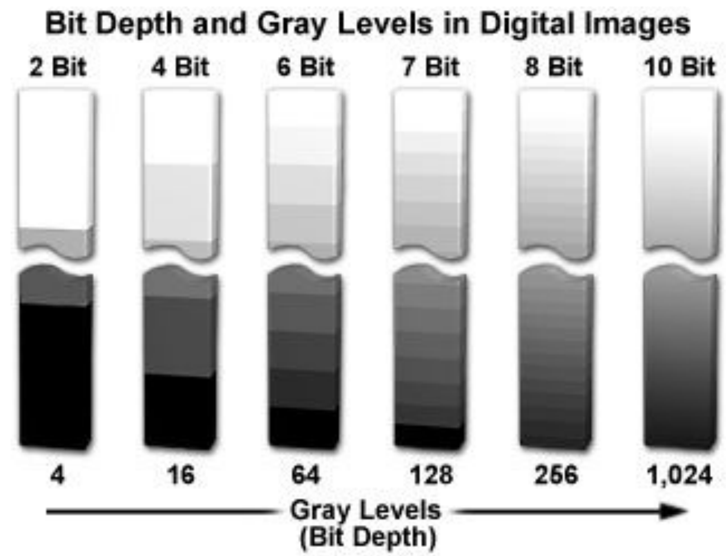
$$\underbrace{6.63}_{Mantissa} \times \underbrace{10^{-34}}_{Exponent}$$

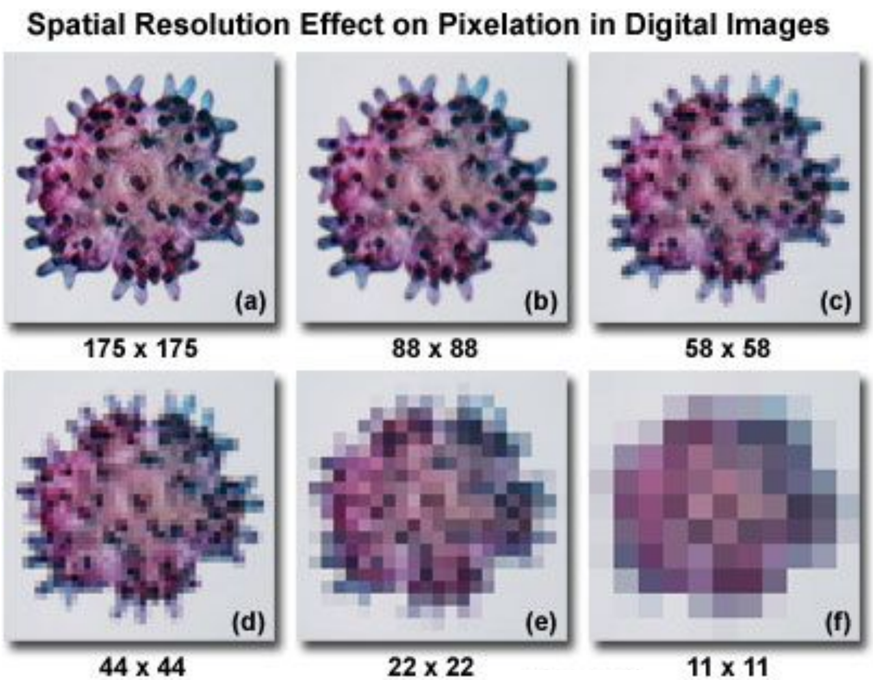


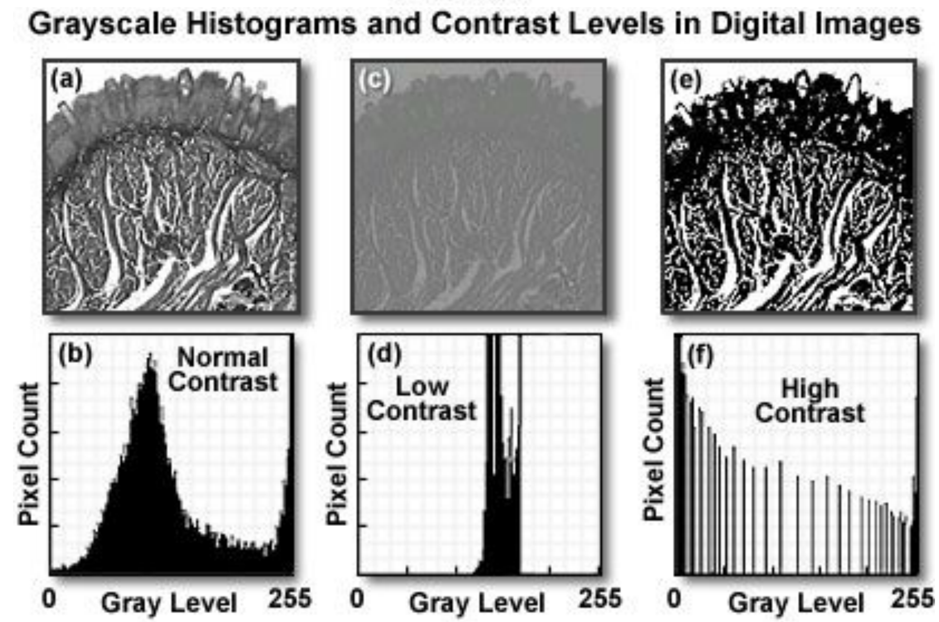
Digital image representation



249	244	240	230	209	233	227	251	255
248	245	210	93	81	120	97	193	254
250	170	133	94	137	120	104	145	253
241	116	118	107	134	138	96	92	163
277	142	121	113	124	115	107	71	179
234	106	84	125	97	108	125	106	204
241	202	102	132	75	73	141	246	252
253	252	244	239	178	199	242	250	245
255	249	244	250	226	231	240	251	253







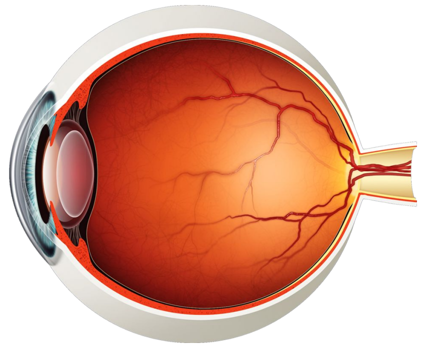
from: http://hamamatsu.magnet.fsu.edu/articles/digitalimagebasics.html

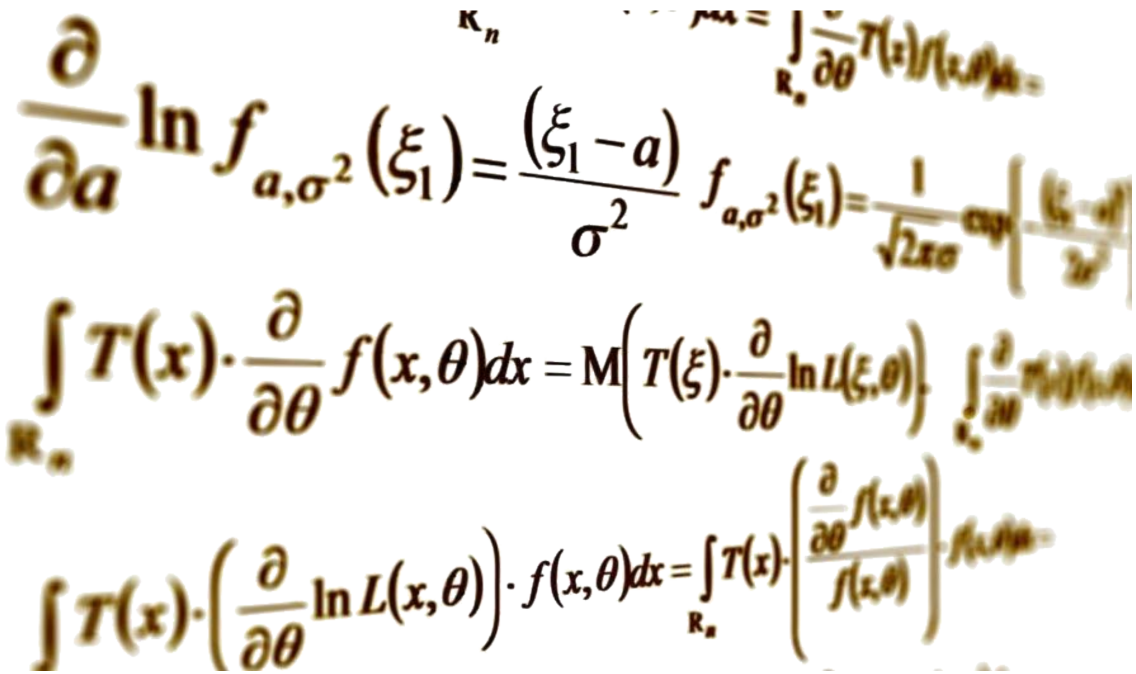
Histograms of pixel values

from: http://demo.ipol.im/demo/27/

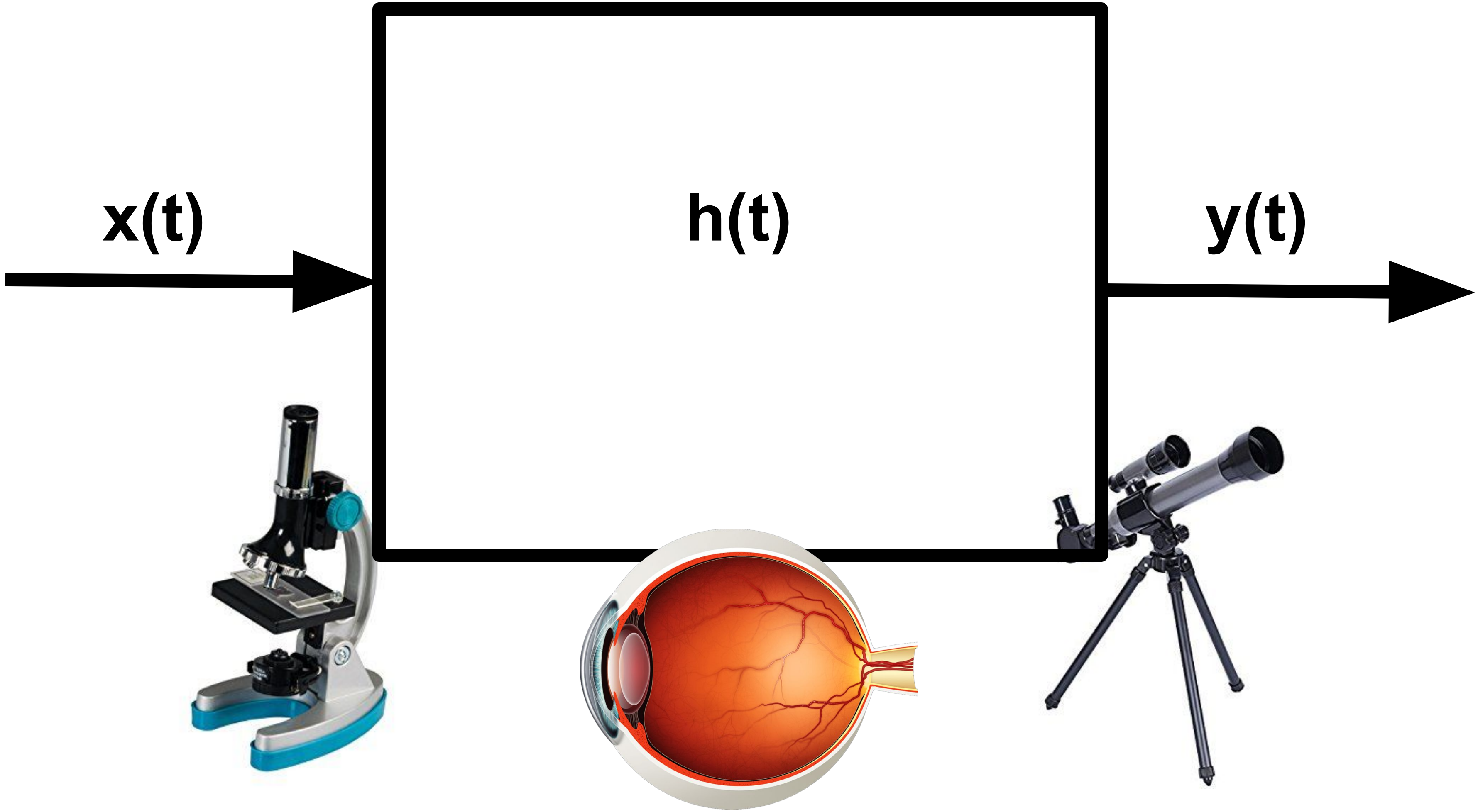
Histograms of pixel values

from: http://demo.ipol.im/demo/27/

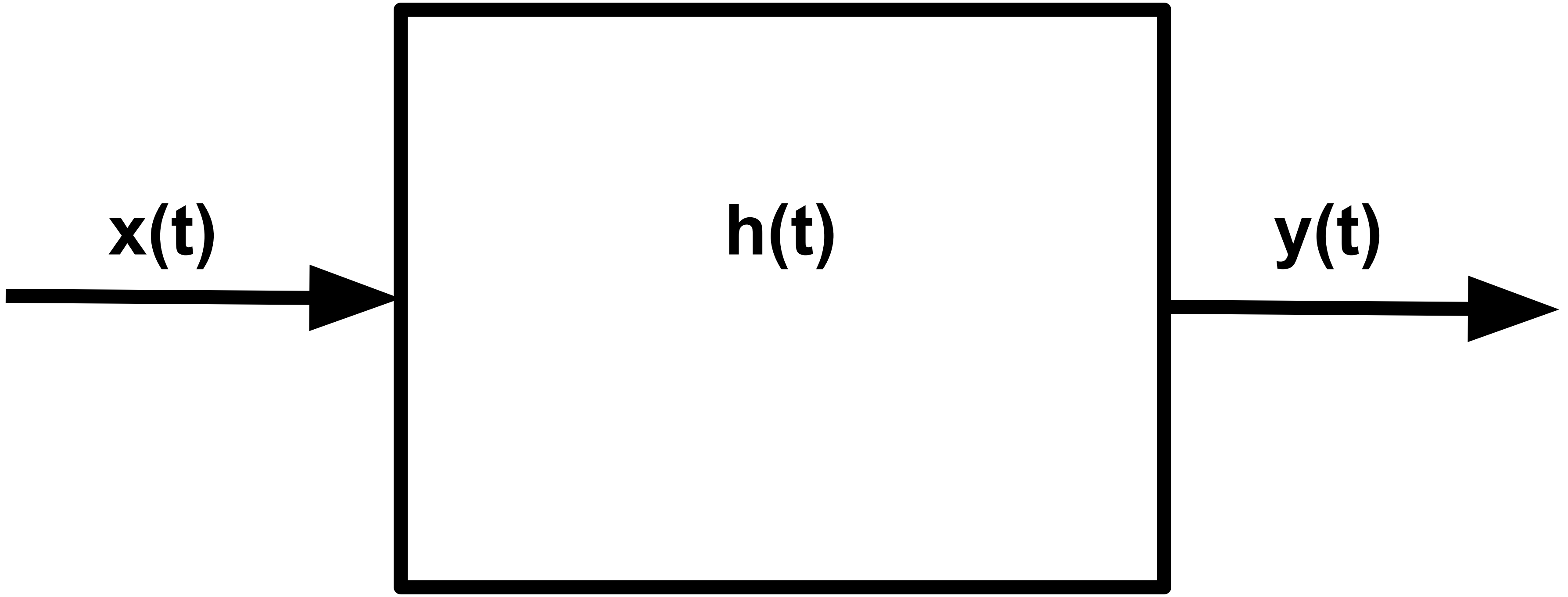



$\frac{\partial}{\partial a}\ln f_{a,\sigma^2}(\xi_1)=\frac{(\xi_1-a)}{\sigma^2}$
$\int_{R_n} T(x)\cdot\frac{\partial}{\partial\theta}f(x,\theta)dx = M\left(T(\xi)\cdot\frac{\partial}{\partial\theta}\ln L(\xi,\theta)\right)$
$\int T(x)\cdot\left(\frac{\partial}{\partial\theta}\ln L(x,\theta)\right)\cdot f(x,\theta)dx$

Linear time-invariant system

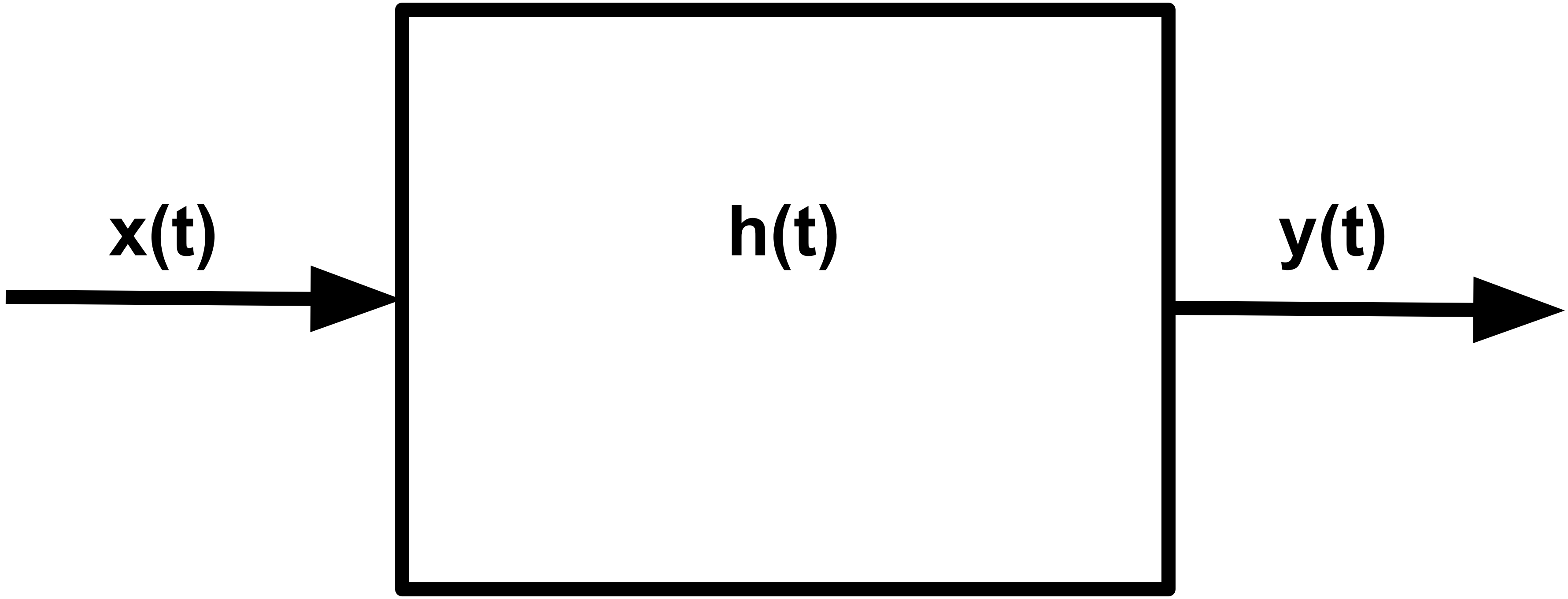


Linear time-invariant system



$$y(t) = x(t) * h(t)$$

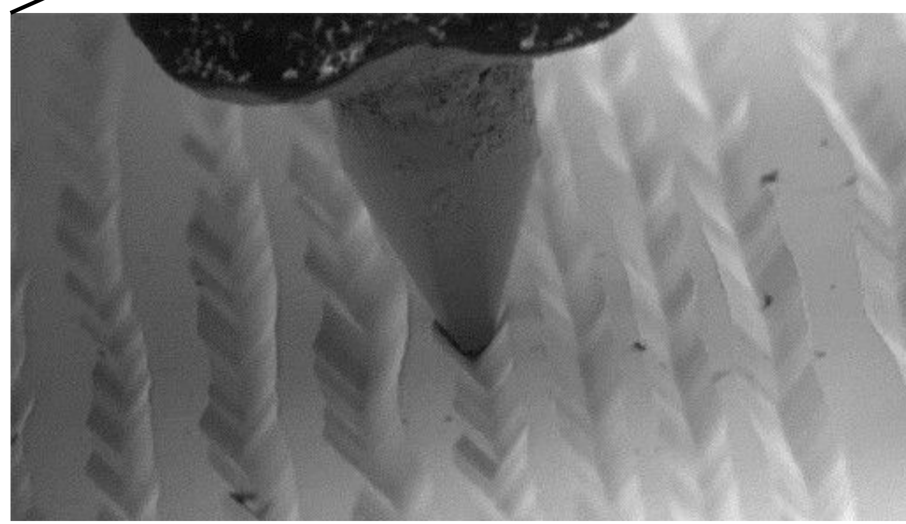
Linear time-invariant system



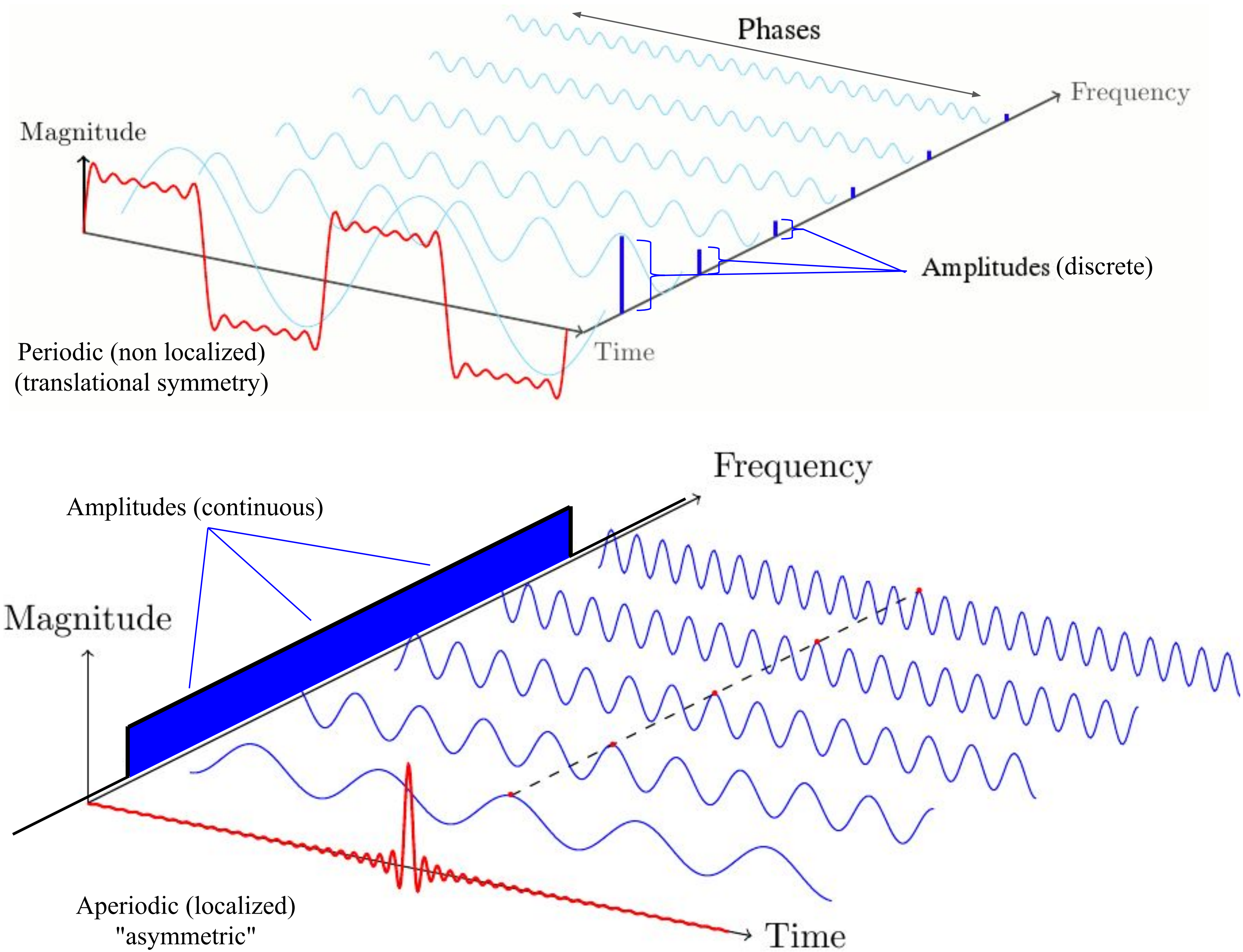
$$y(t) = x(t)*h(t)$$

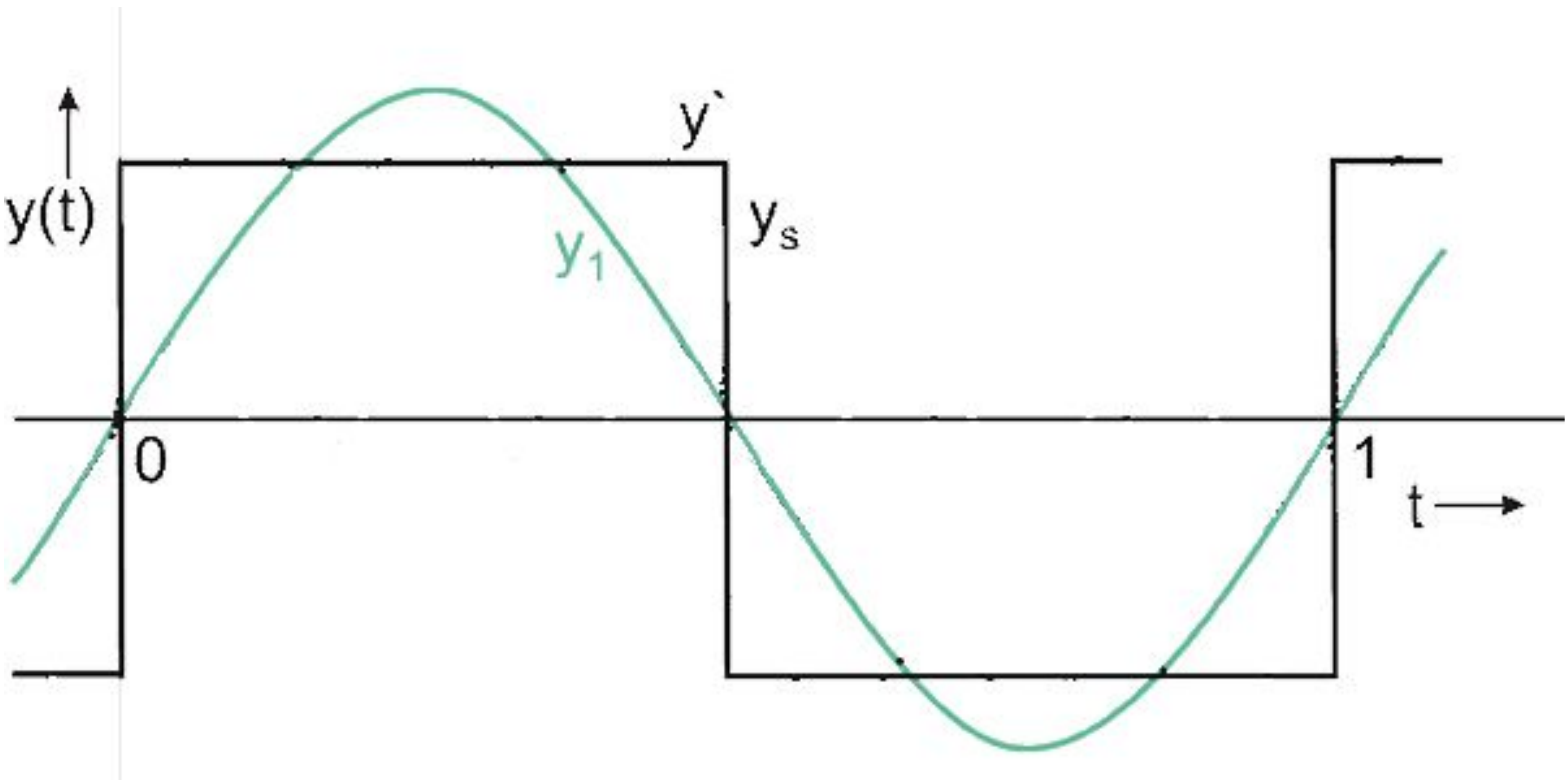
CONVOLUTION

Marcia
"Oregon, My Oregon"
Land of the
Land of the
Empire Builders, Land of the Gold-en West; Con-quered and held by free men,
rose and sunshine, Land of the sum-mer's breeze; Lad-en with health and vig-or,
Fair-est and the best. On-ward and up-ward ev-er, Forward and on, and
Fresh from the Western seas. Blest by the blood of mar-tyrs, Land of the set-ting
on; Hail to thee, Land of He-roes, My O-re-gon.
sun; Hail to thee, Land of Prom-ise, My O-re-gon.

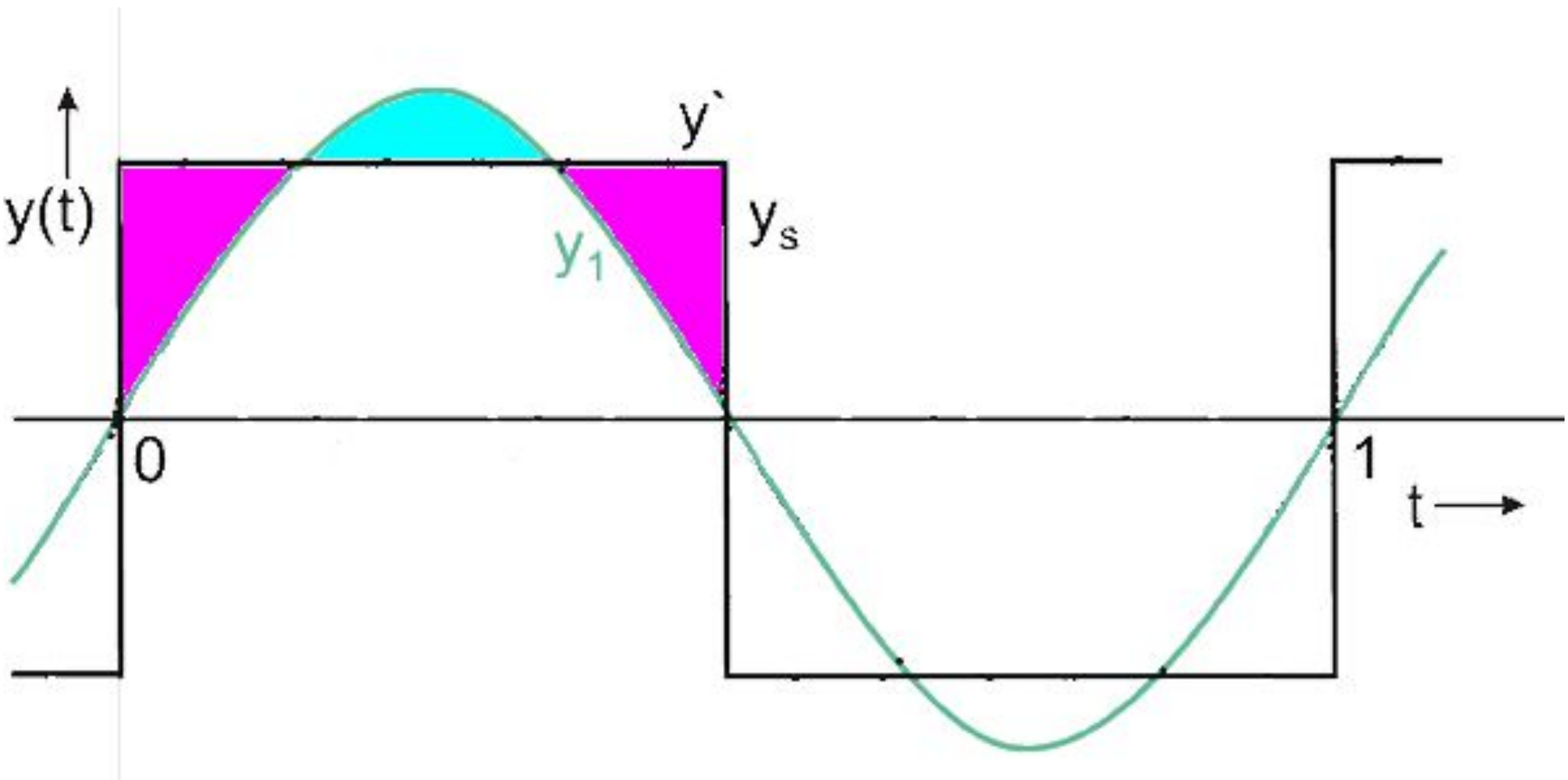


SPLOID GIF

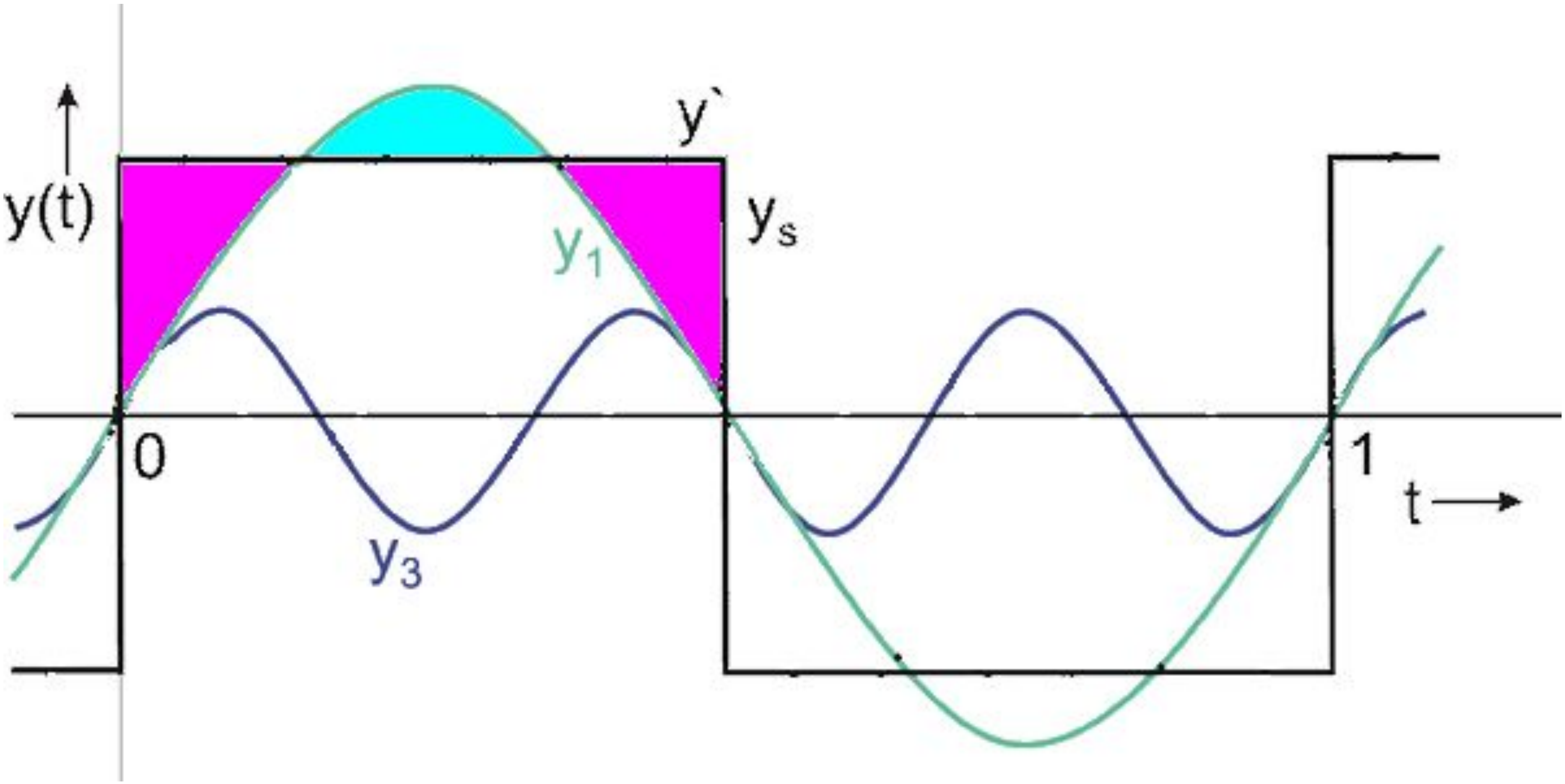

Phases
Frequency
Magnitude
Amplitudes (discrete)
Time
Periodic (non localized)
(translational symmetry)
Frequency
Amplitudes (continuous)
Magnitude
Time
Aperiodic (localized)
"asymmetric"



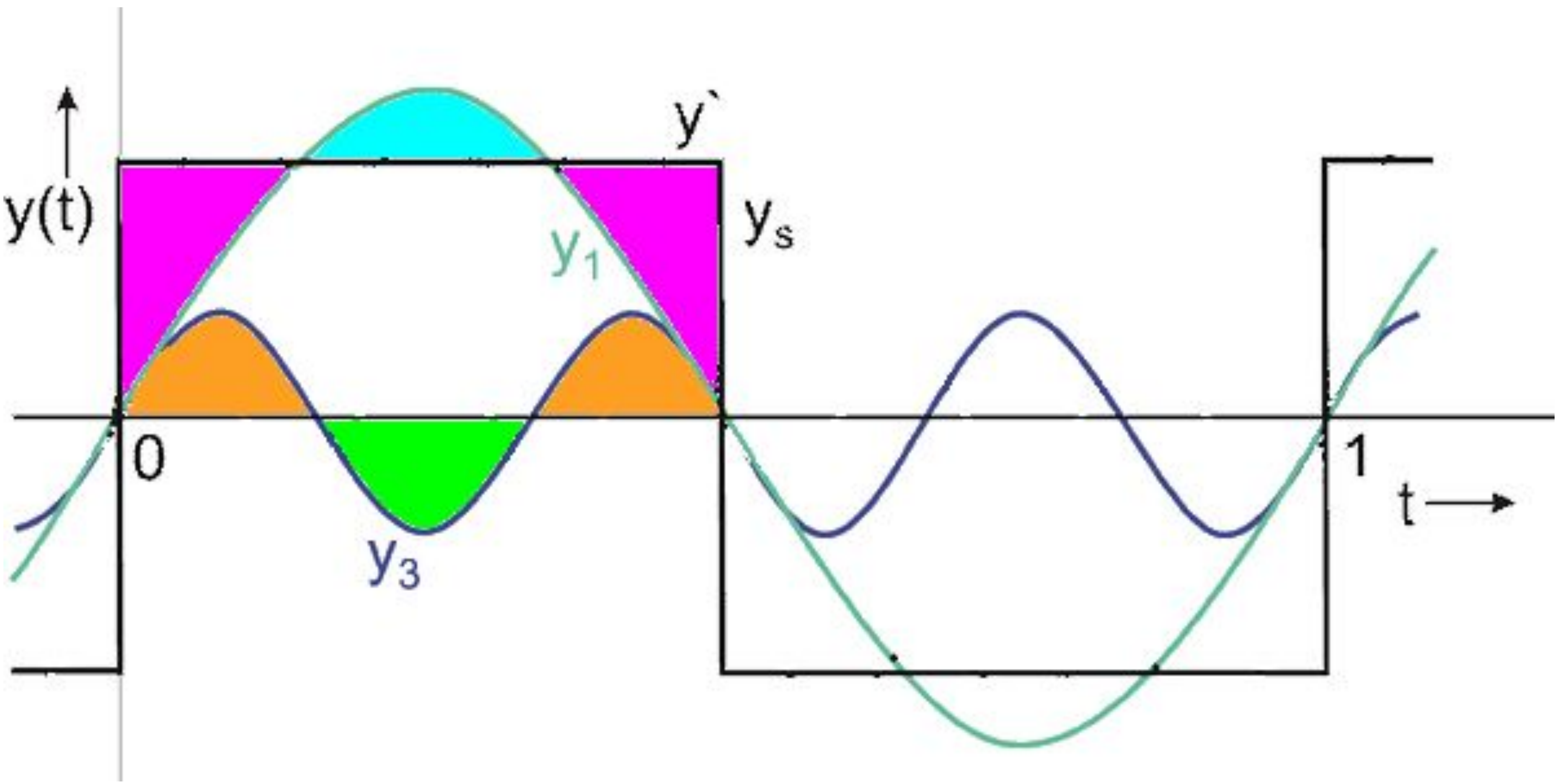
y`
y(t)
y_s
y_1
0
1
t



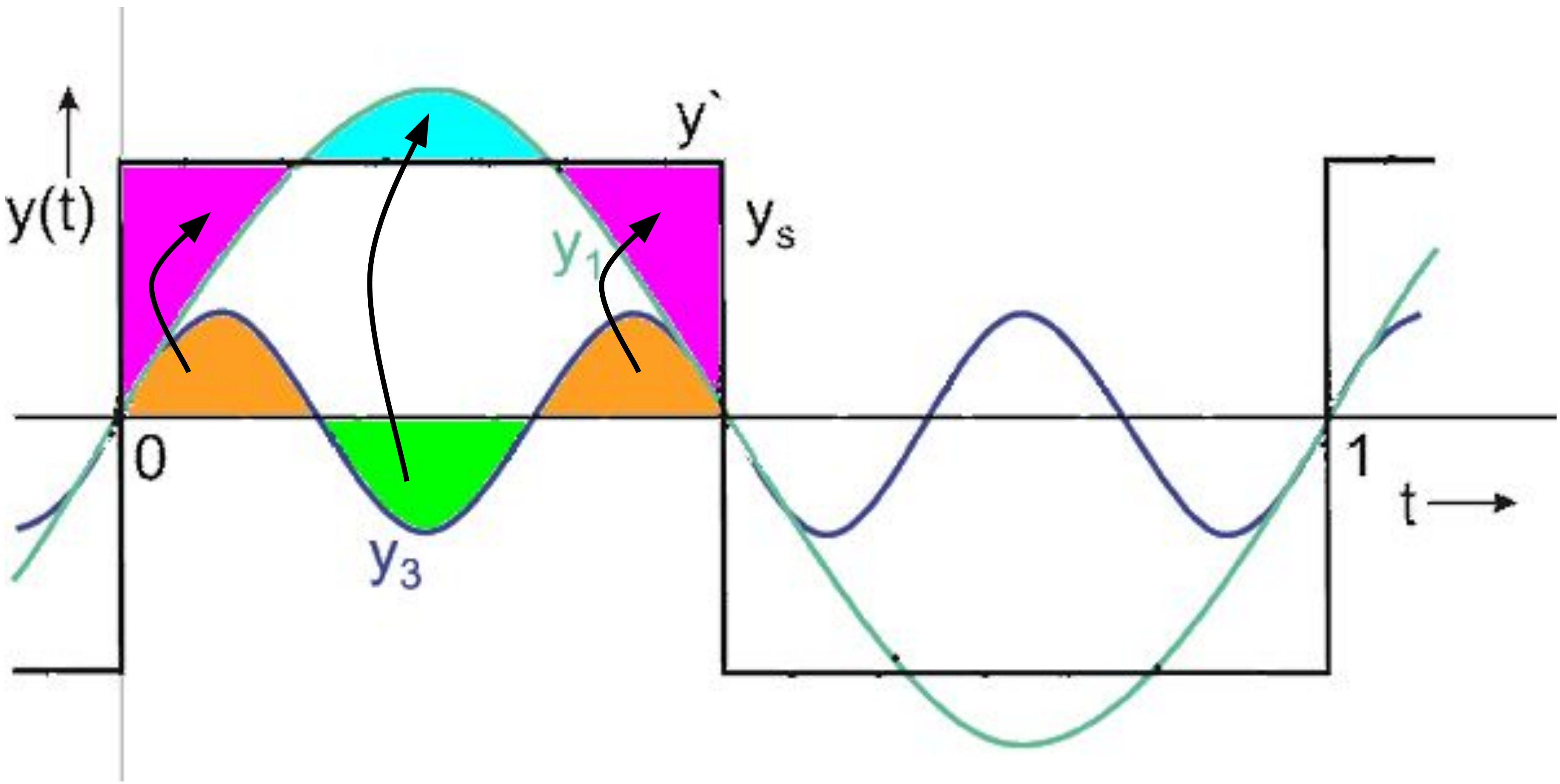
y(t)
y`
y_s
y_1
0
1
t

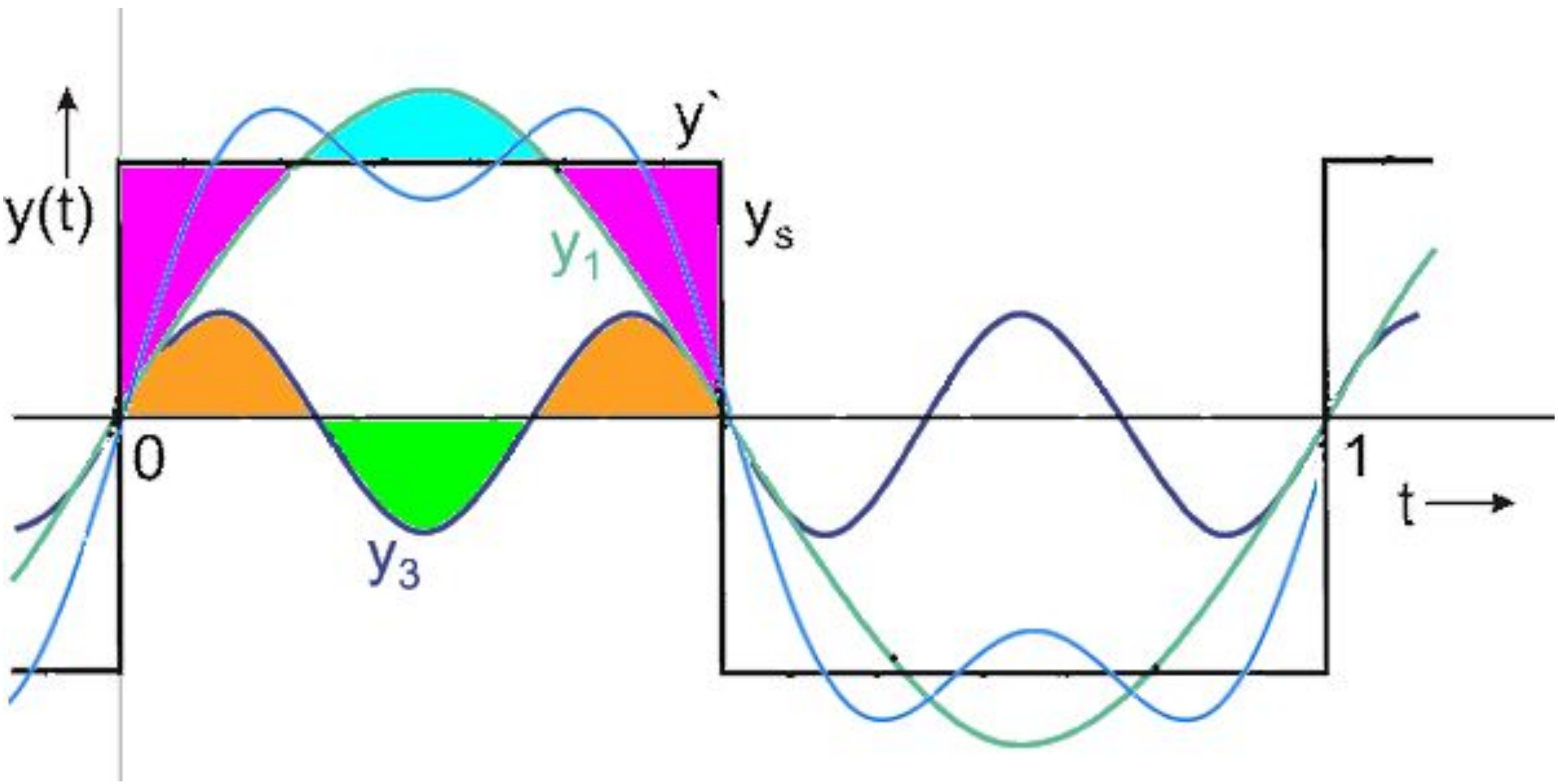


y(t)
y`
y_s
y_1
y_3
0
1
t

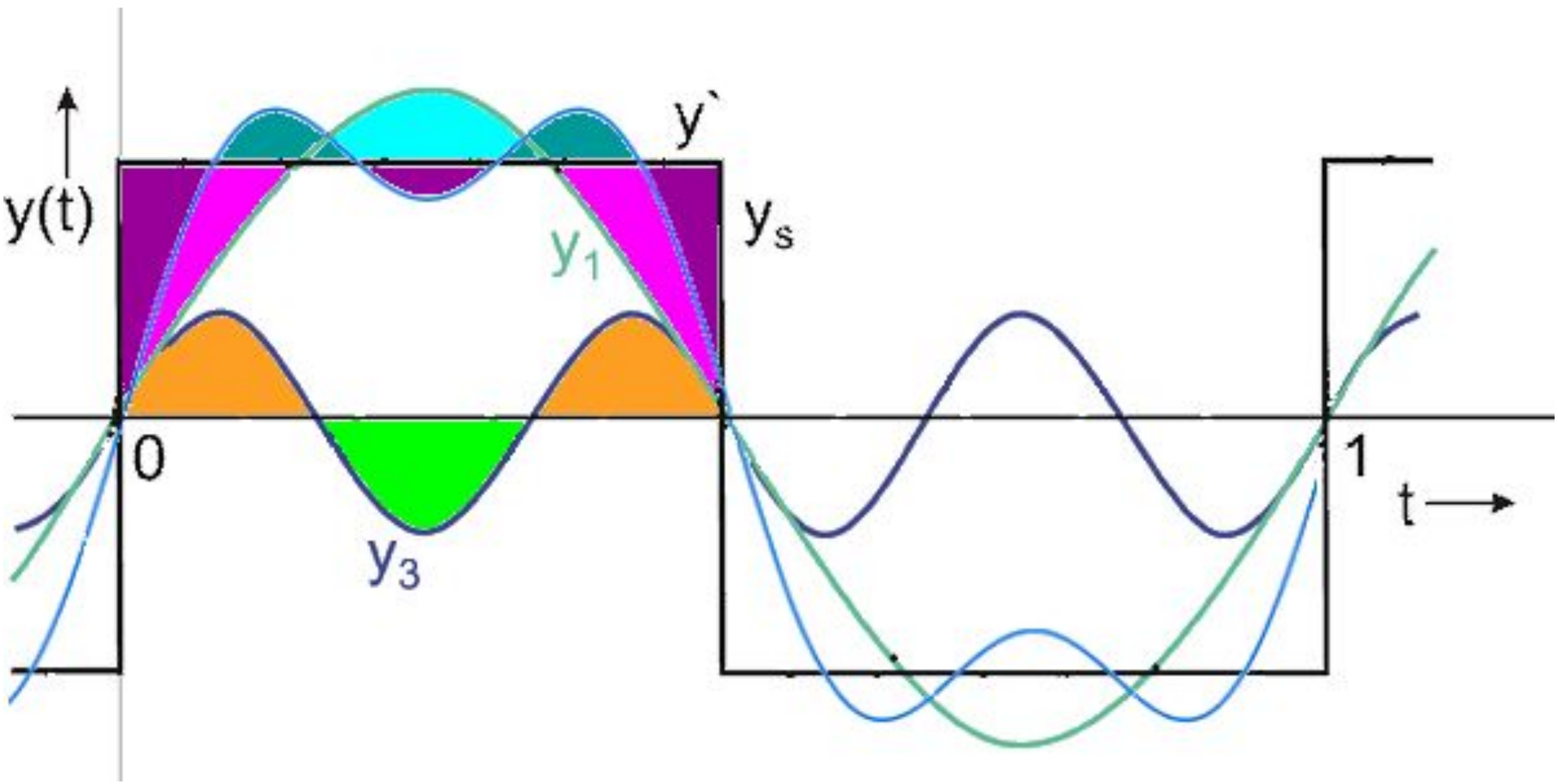


y(t)
y`
y_s
y_1
y_3
0
1
t

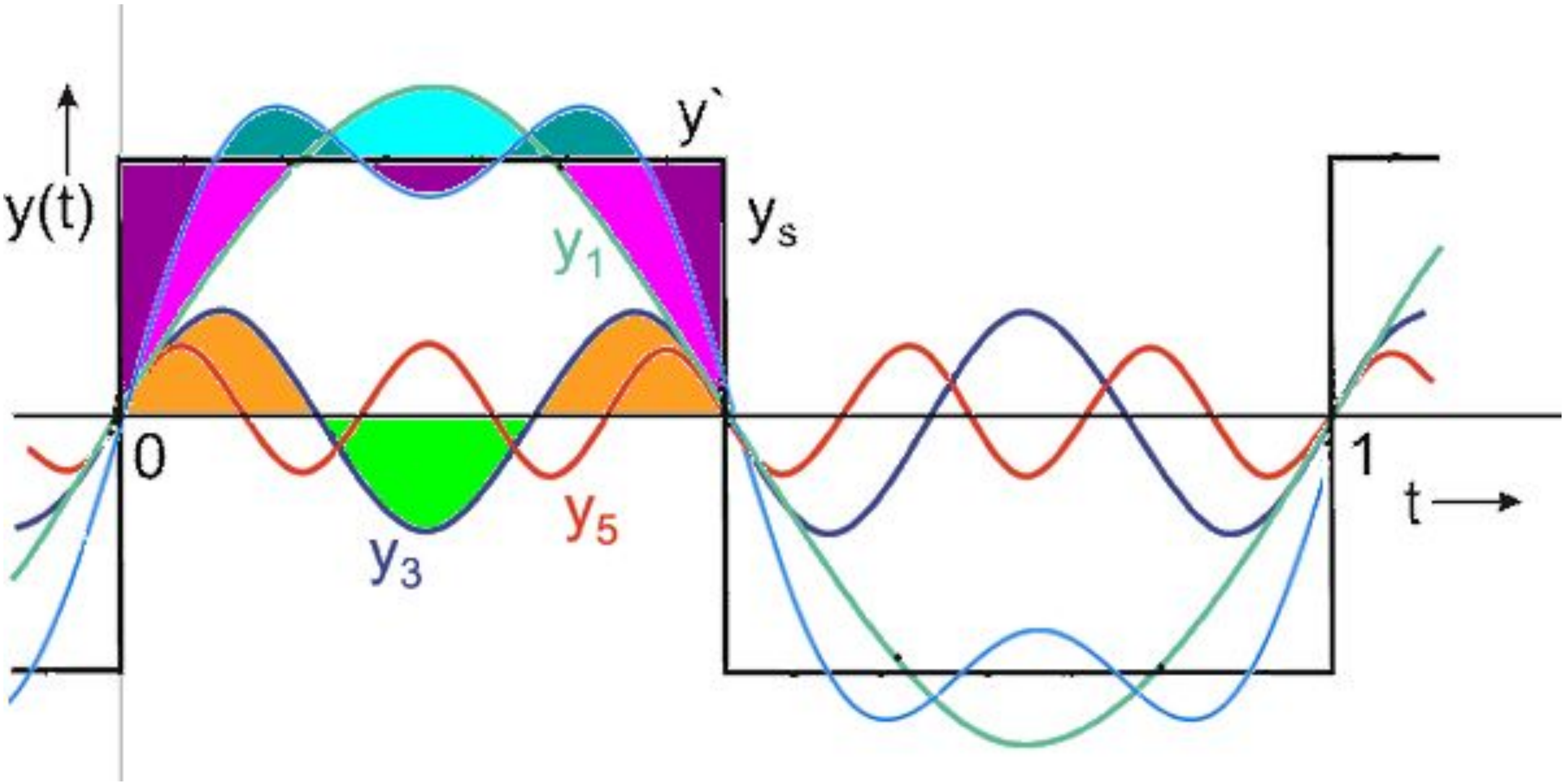

y(t)
y`
y_s
y_1
y_3
0
1
t



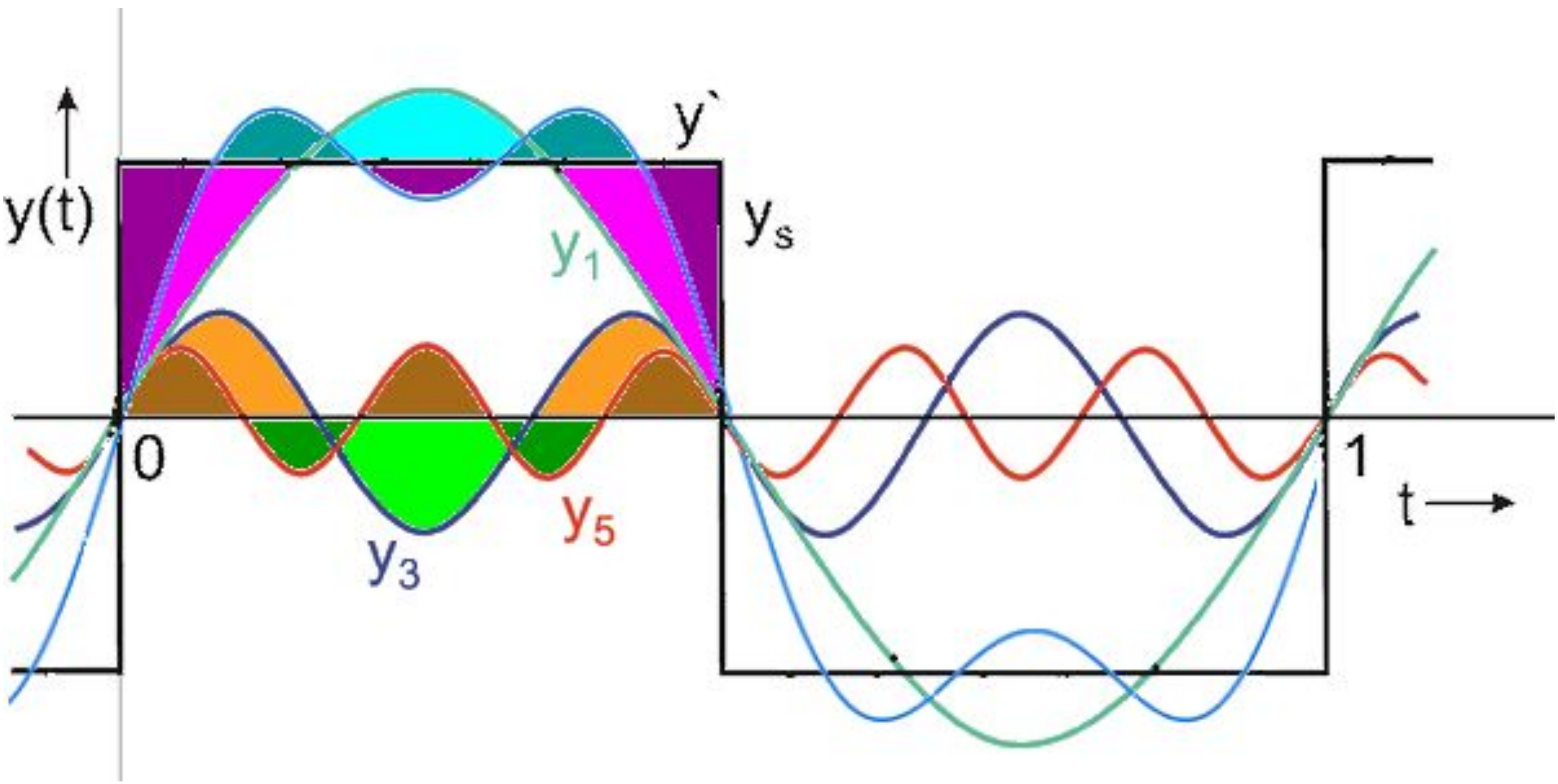
y(t)
y`
y_s
y_1
y_3
0
1
t

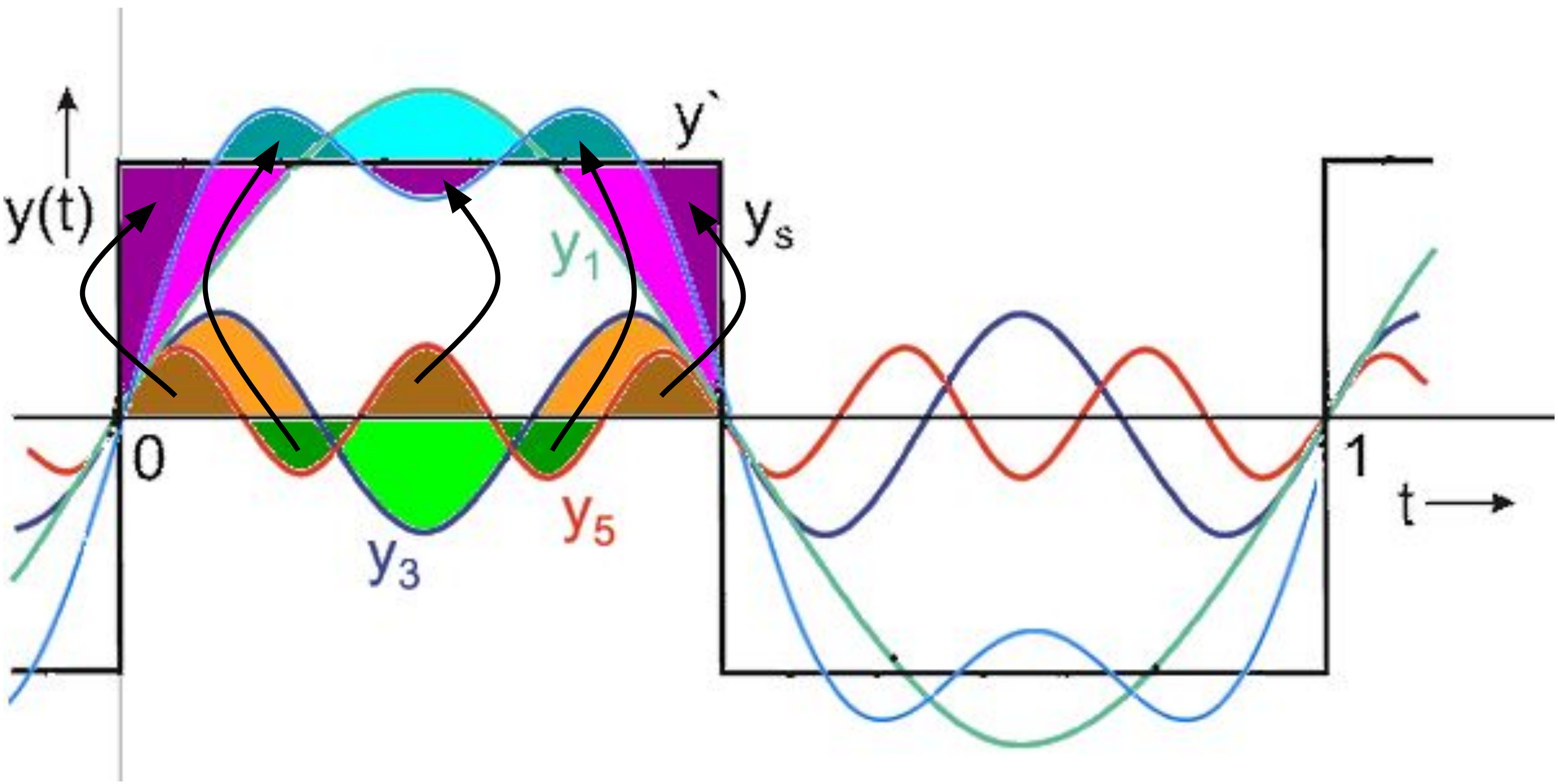


y(t)
y`
y_s
y_1
y_3
0
1
t

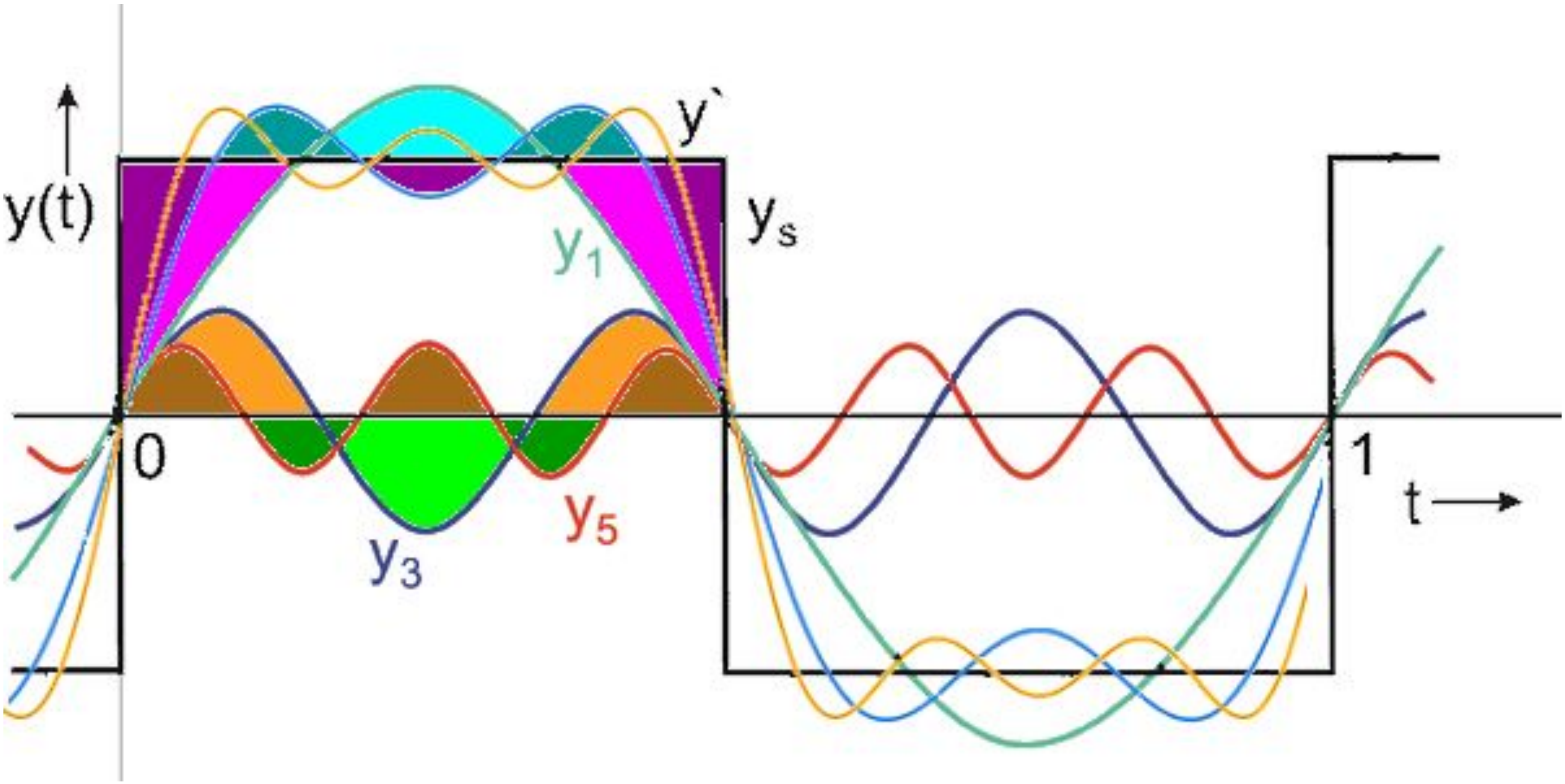


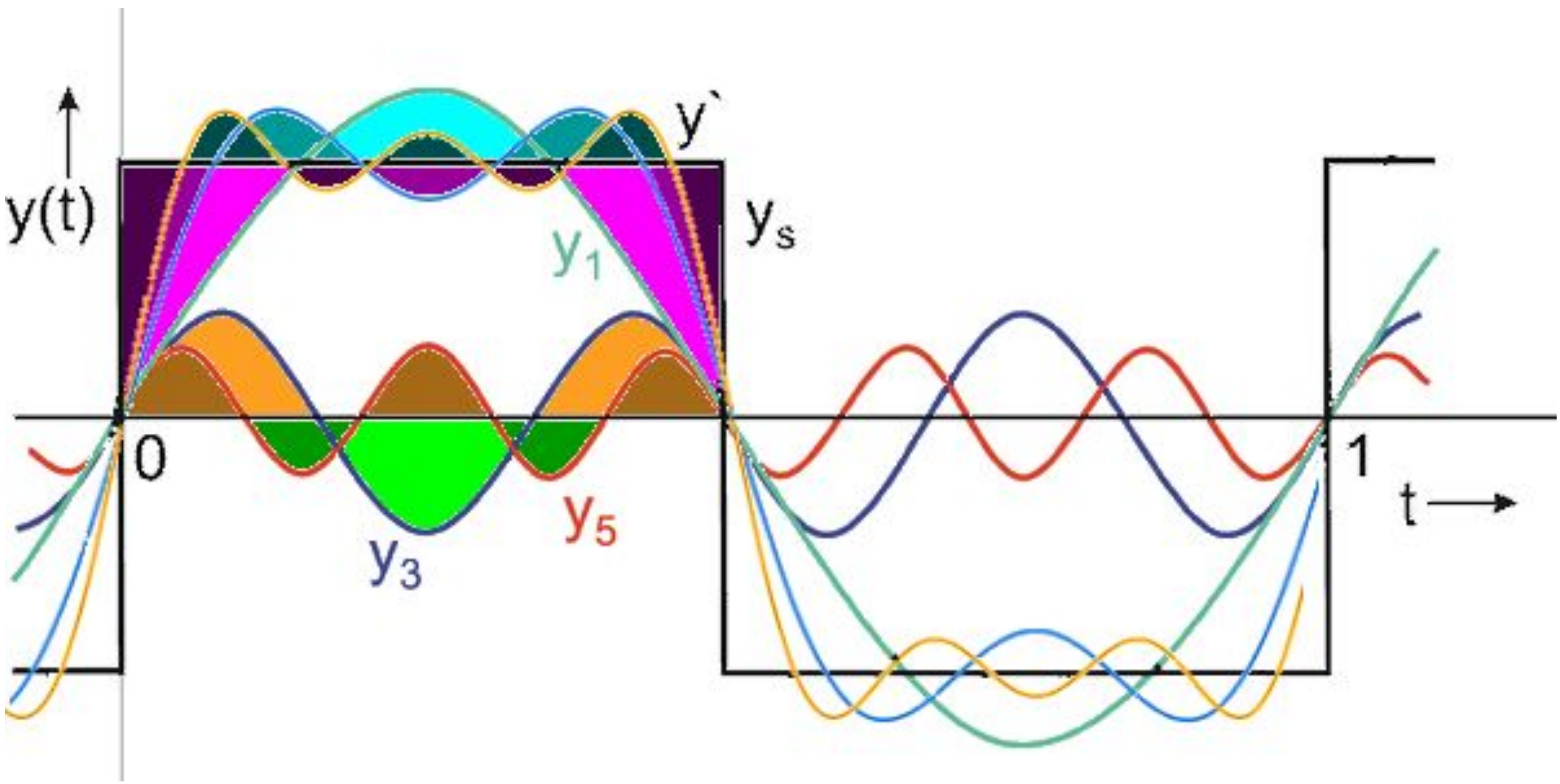
y(t)
y`
y_s
y_1
y_3
y_5
0
1
t


y(t)
y`
y_s
y_1
y_3
y_5
0
1
t

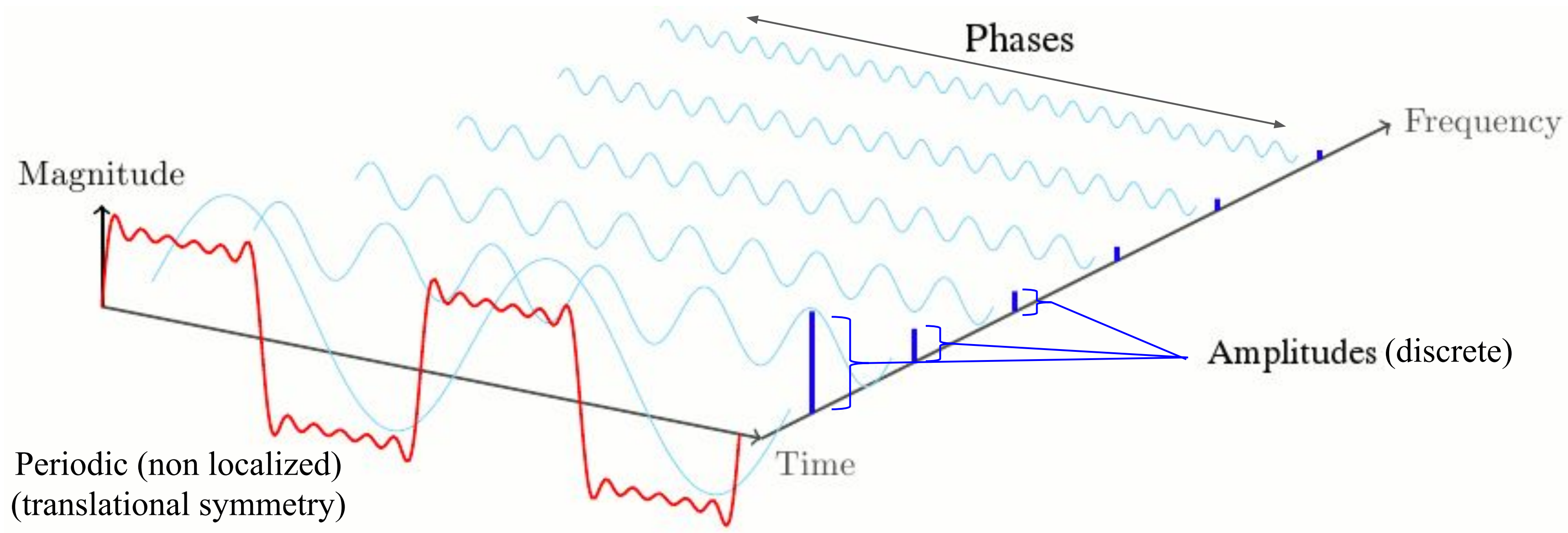


y(t)
y`
y_s
y_1
0
1
t
y_3
y_5

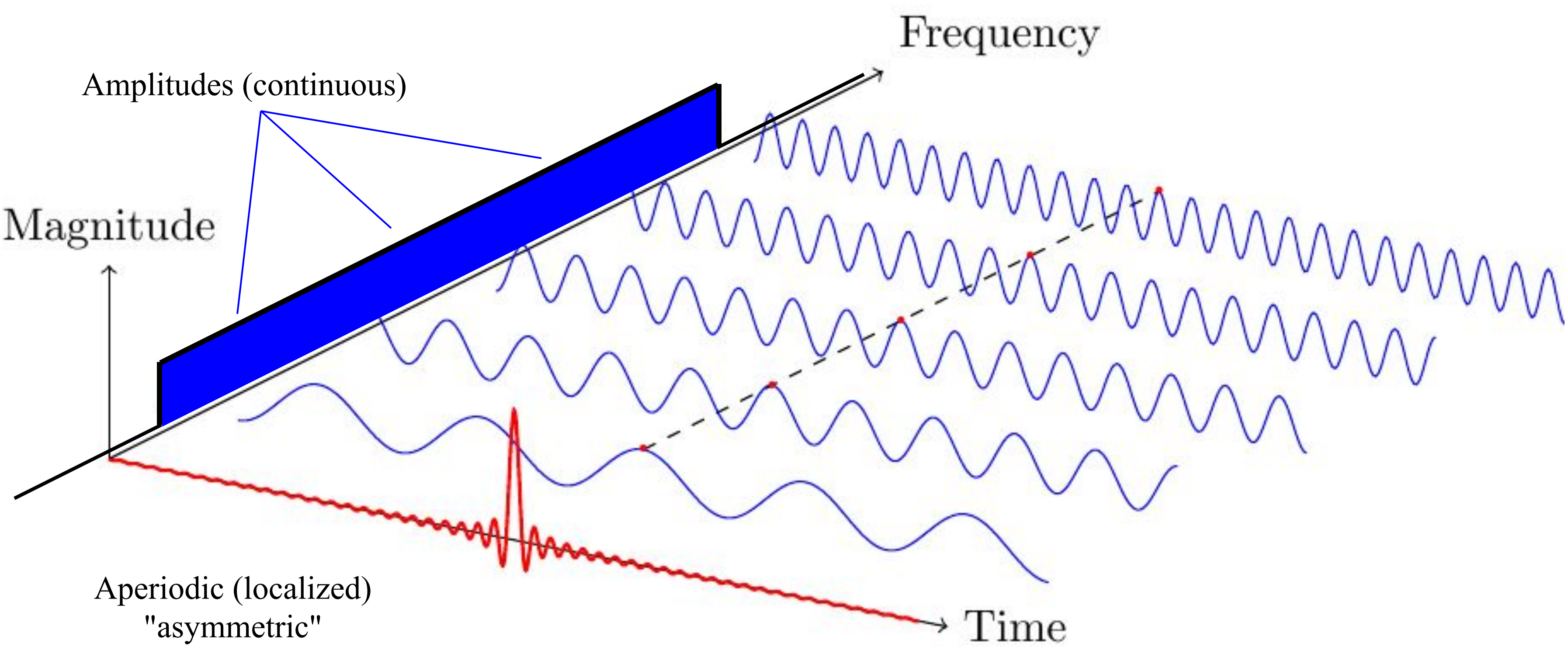

y(t)
y`
ys
y1
0
1
t
y5
y3



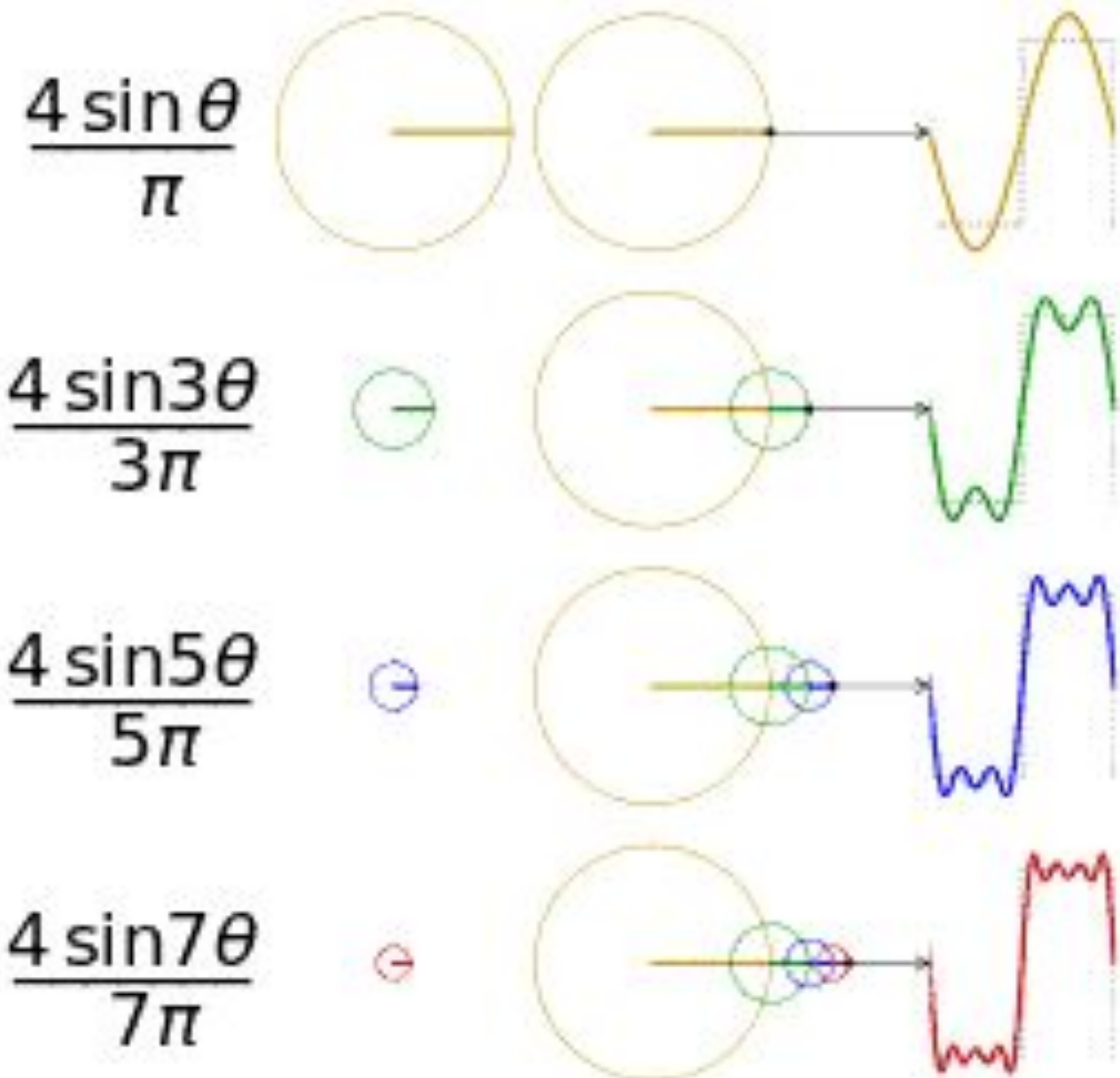
y(t)
y`
ys
y1
0
y3
y5
1
t



Phases
Frequency
Magnitude
Amplitudes (discrete)
Time
Periodic (non localized)
(translational symmetry)



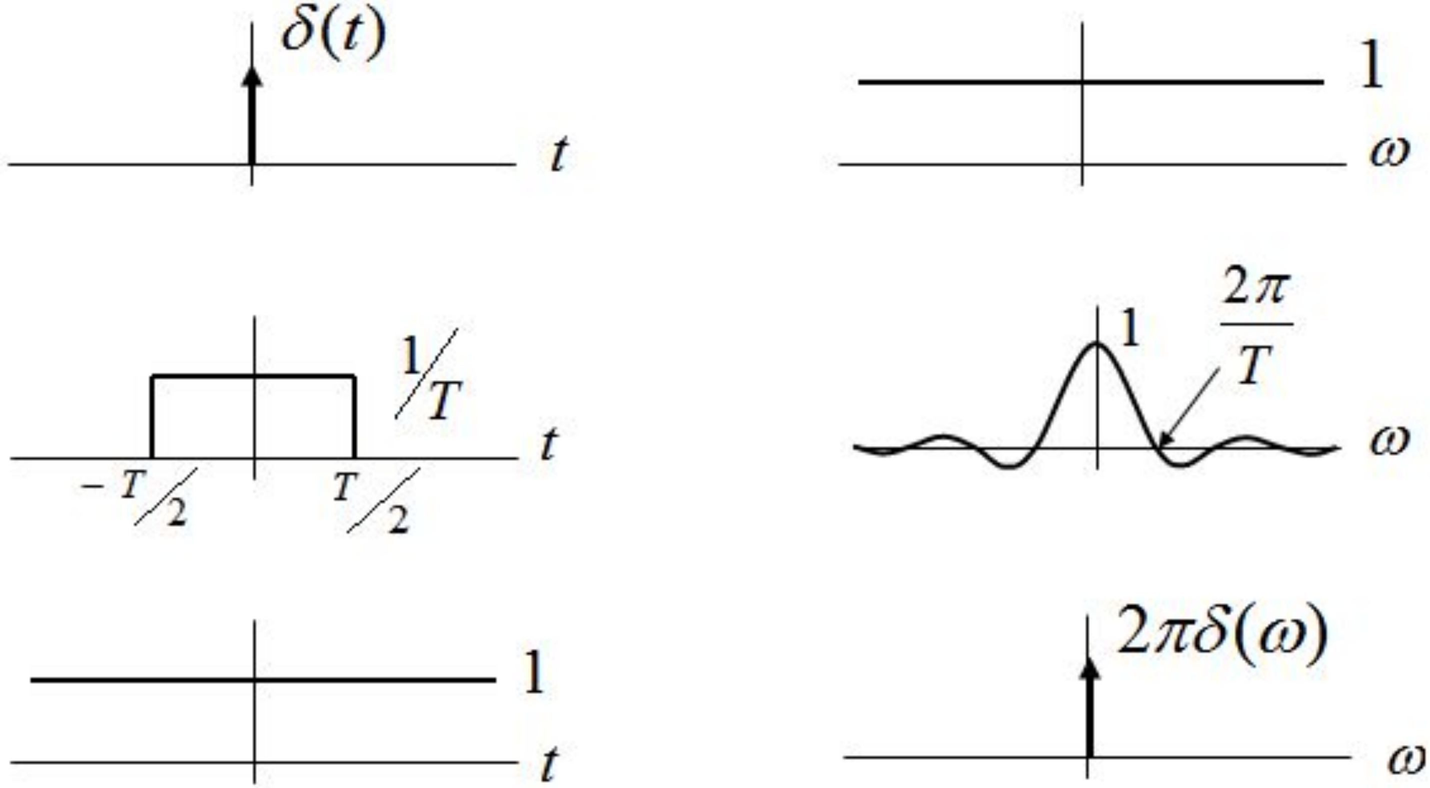
Frequency
Amplitudes (continuous)
Magnitude
Aperiodic (localized)
"asymmetric"
Time



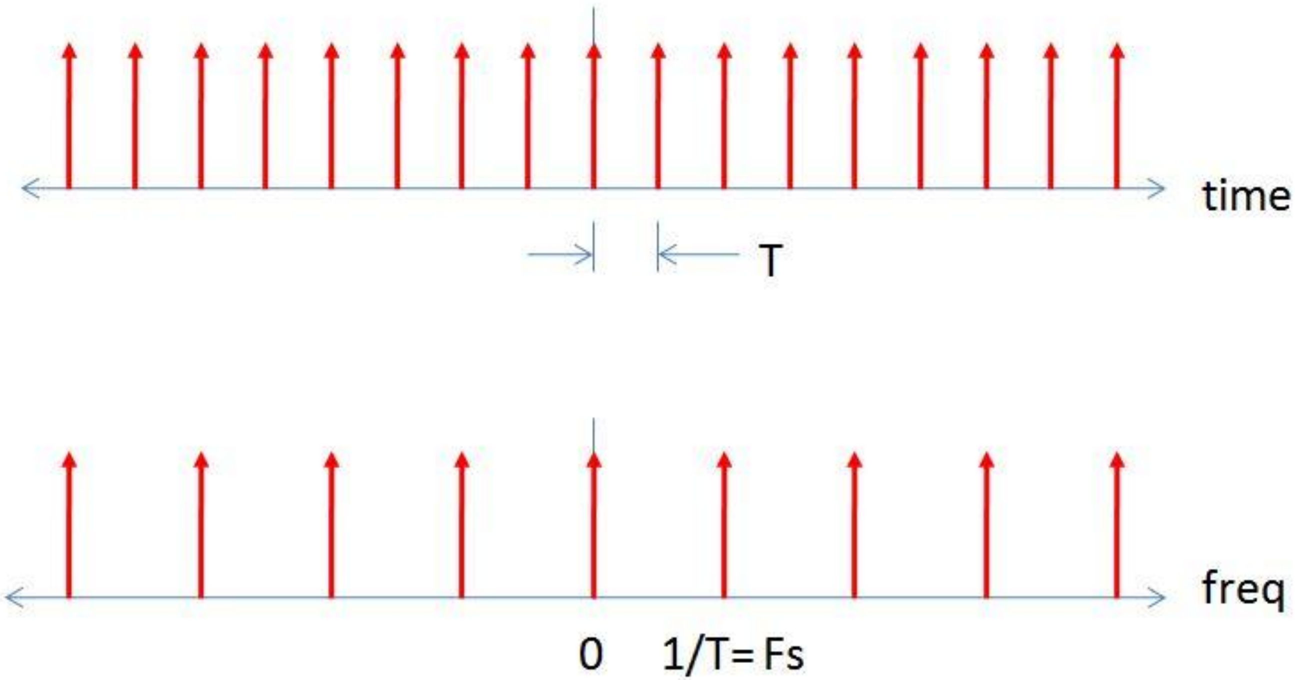
From: https://en.wikipedia.org/wiki/Fourier_series

Important Fourier Transforms

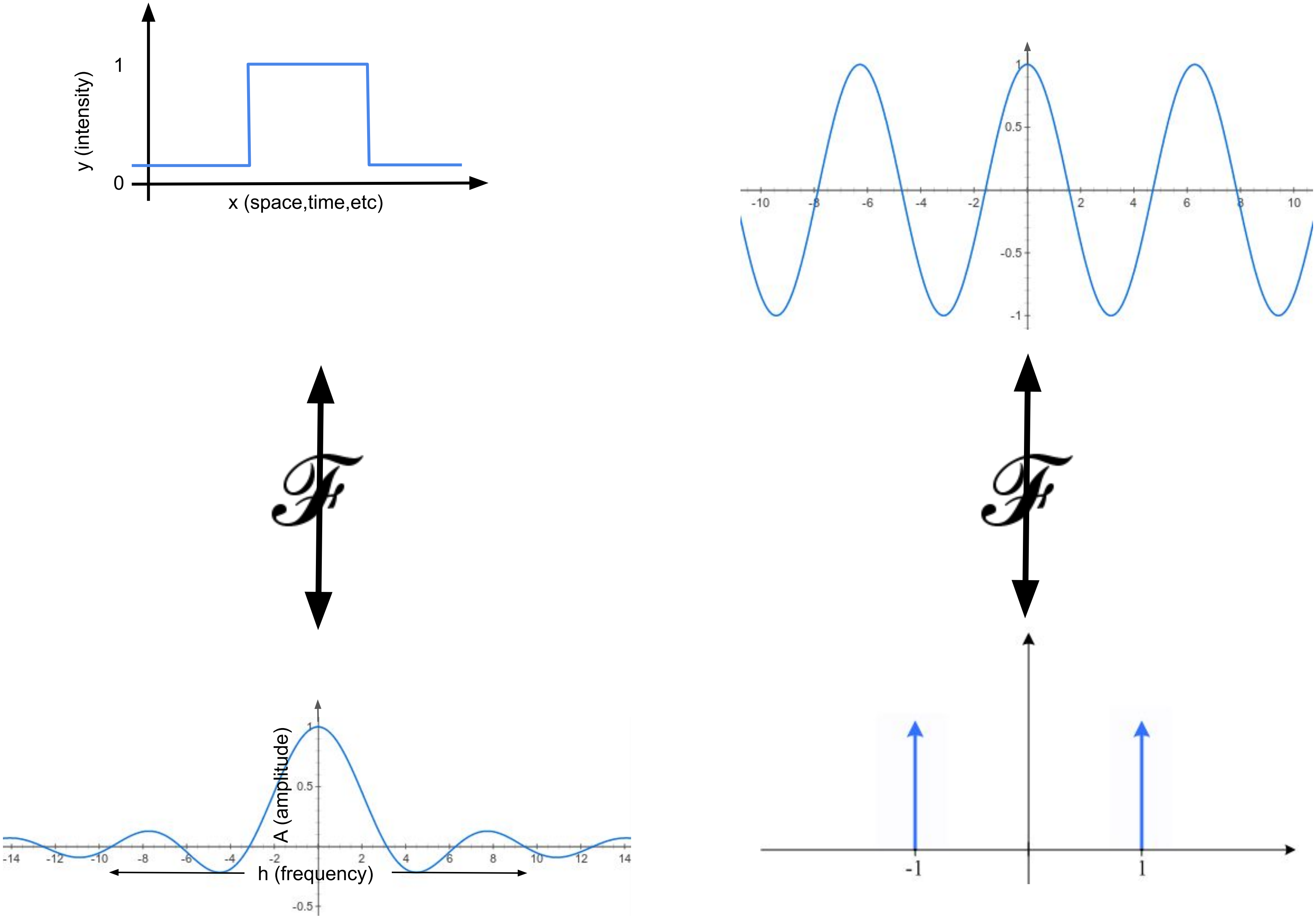
Time Domain | Frequency Domain



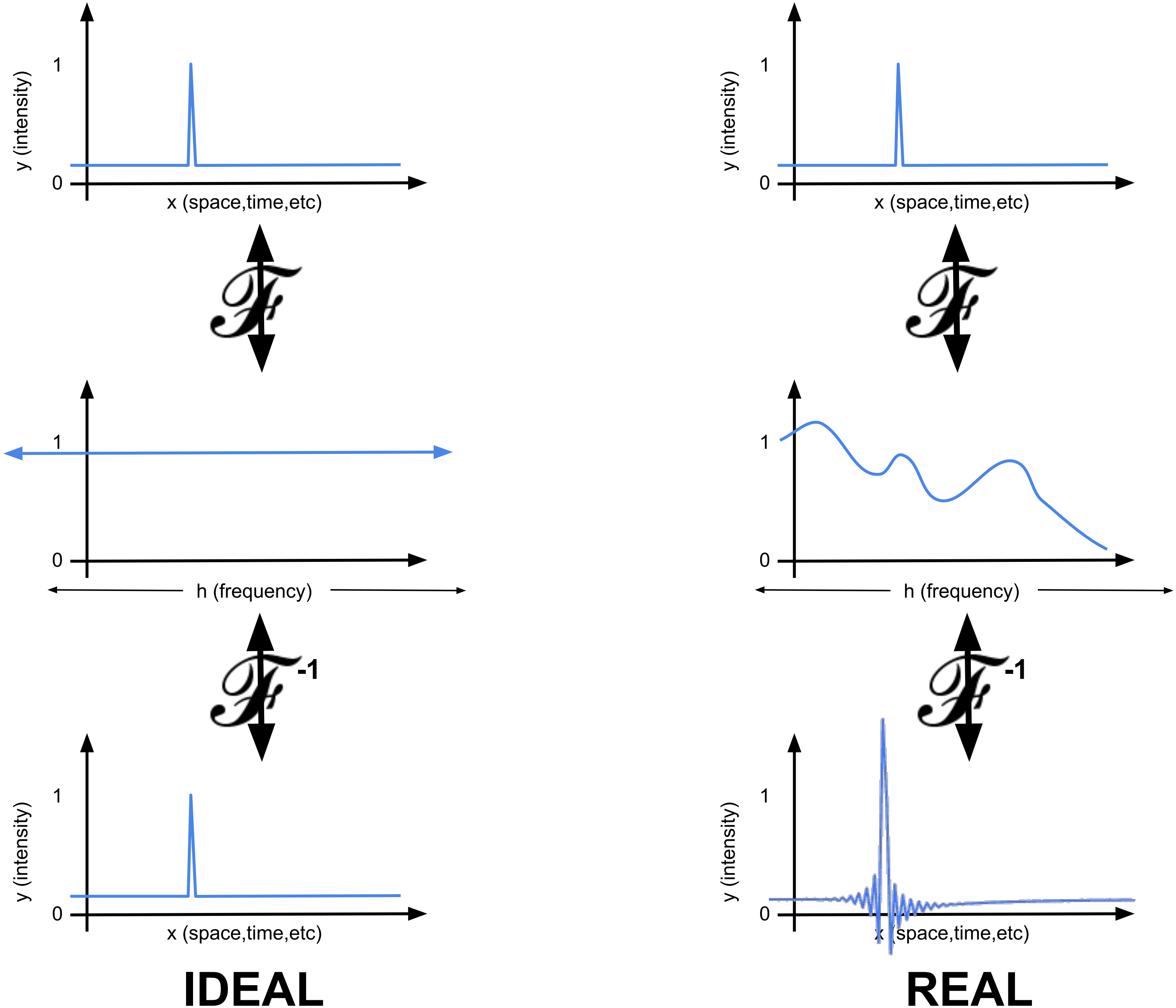
Review – FT of Impulse Train



Impulses in time $\longrightarrow \mathfrak{F}\{\} \longrightarrow$ Impulses in frequency

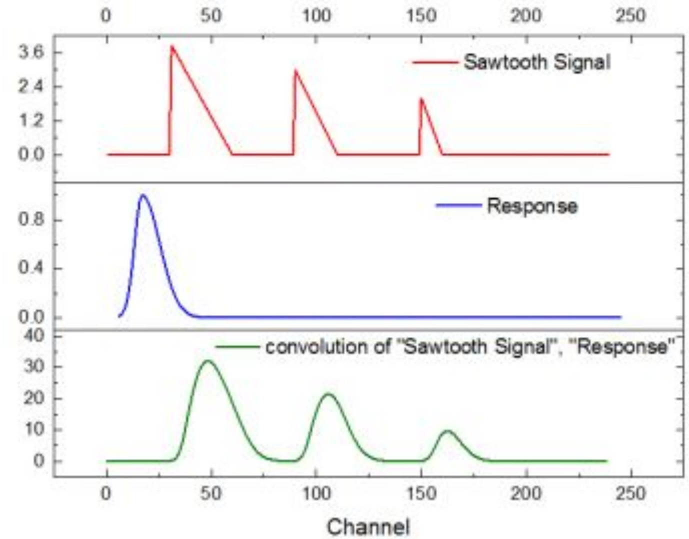


y (intensity)
1
0
x (space,time,etc)
A (amplitude)
h (frequency)
-1
1



y (intensity)
1
0
x (space,time,etc)
-1
h (frequency)
IDEAL
REAL

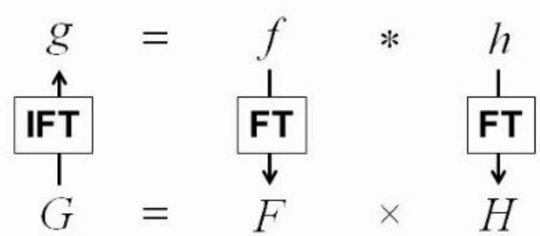
$$(f * g)(t) \stackrel{\mathrm{def}}{=} \int_{-\infty}^{\infty} f(\tau) g(t-\tau)\, d\tau$$
$$= \int_{-\infty}^{\infty} f(t-\tau) g(\tau)\, d\tau.$$

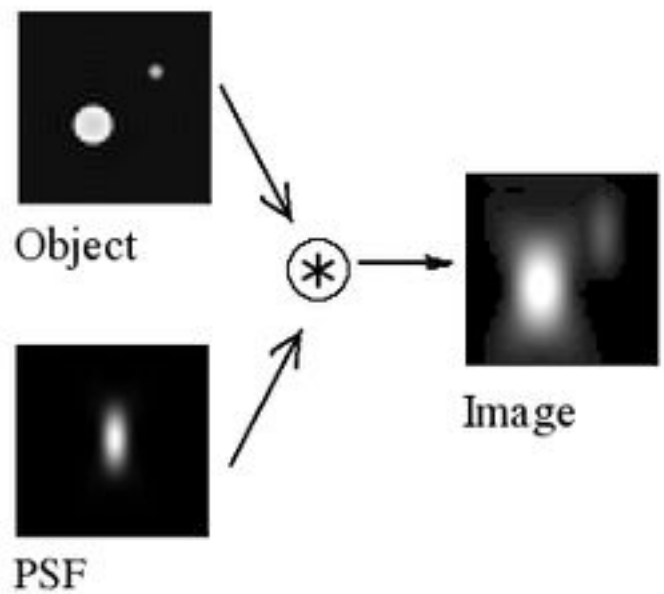


From: https://www.originlab.com/doc/Origin-Help/Convolution

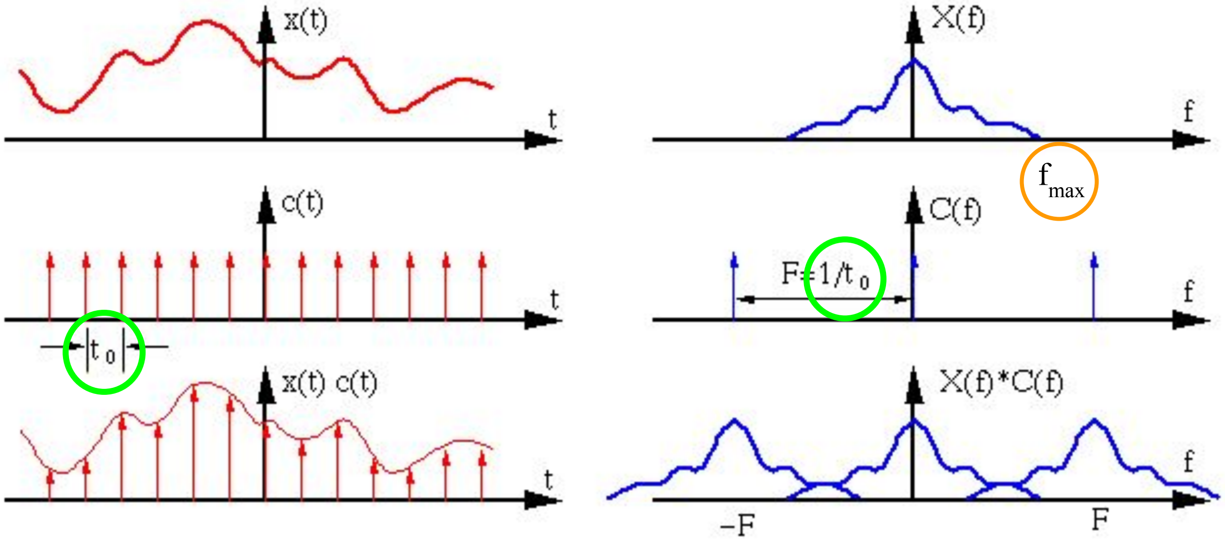
Spatial Domain *(x)*		**Frequency Domain *(u)***
$g = f * h$	$\leftrightarrow$	$G = FH$
$g = fh$	$\leftrightarrow$	$G = F * H$

So, we can find *g(x)* by Fourier transform

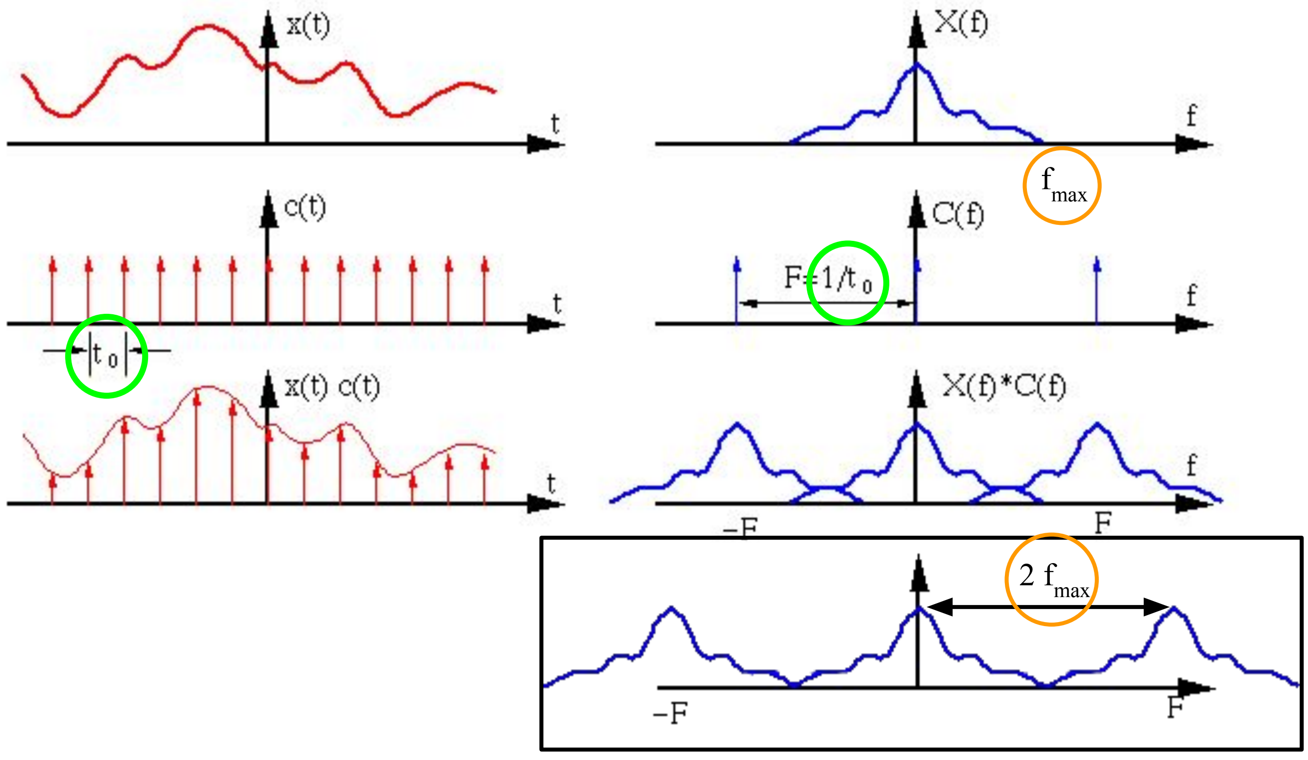




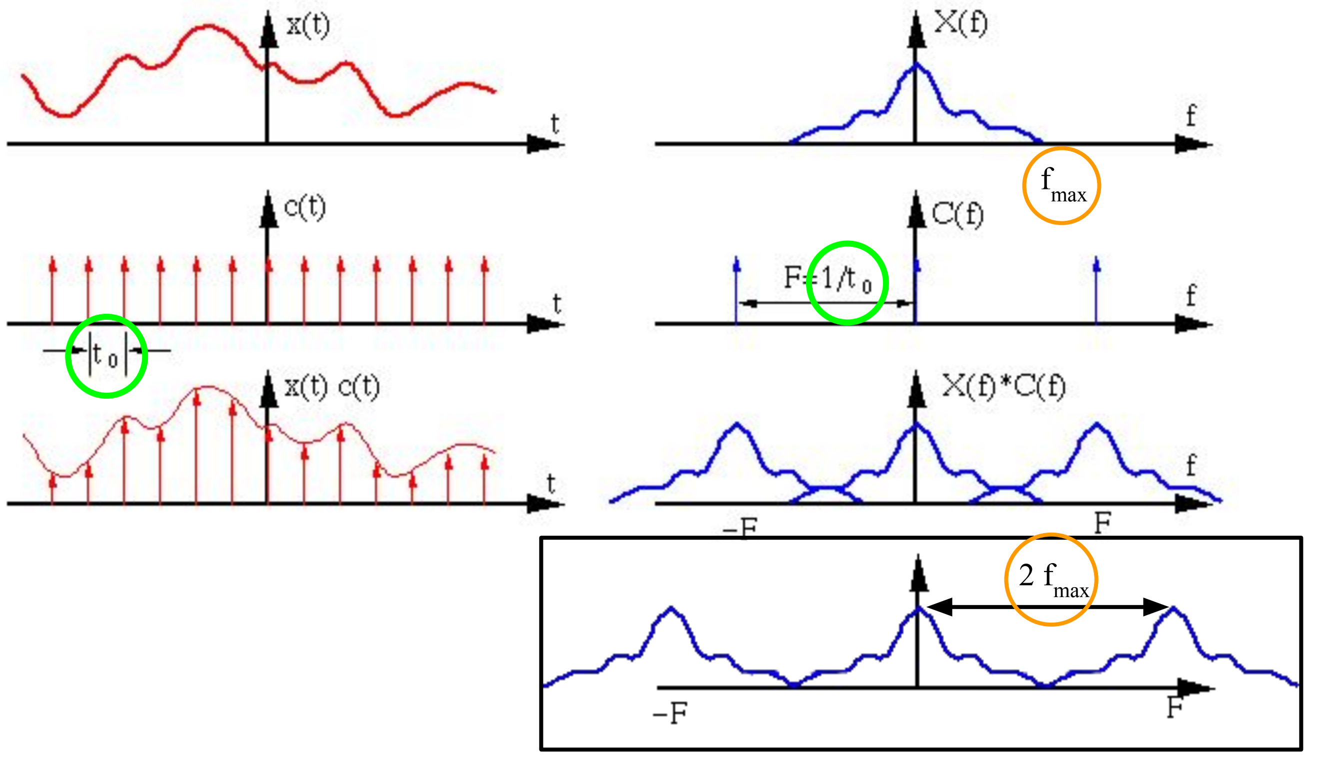
from: https://en.wikipedia.org/wiki/Point_spread_function



from: http://fourier.eng.hmc.edu/e101/lectures/Image_Processing/node2.html



from: http://fourier.eng.hmc.edu/e101/lectures/Image_Processing/node2.html

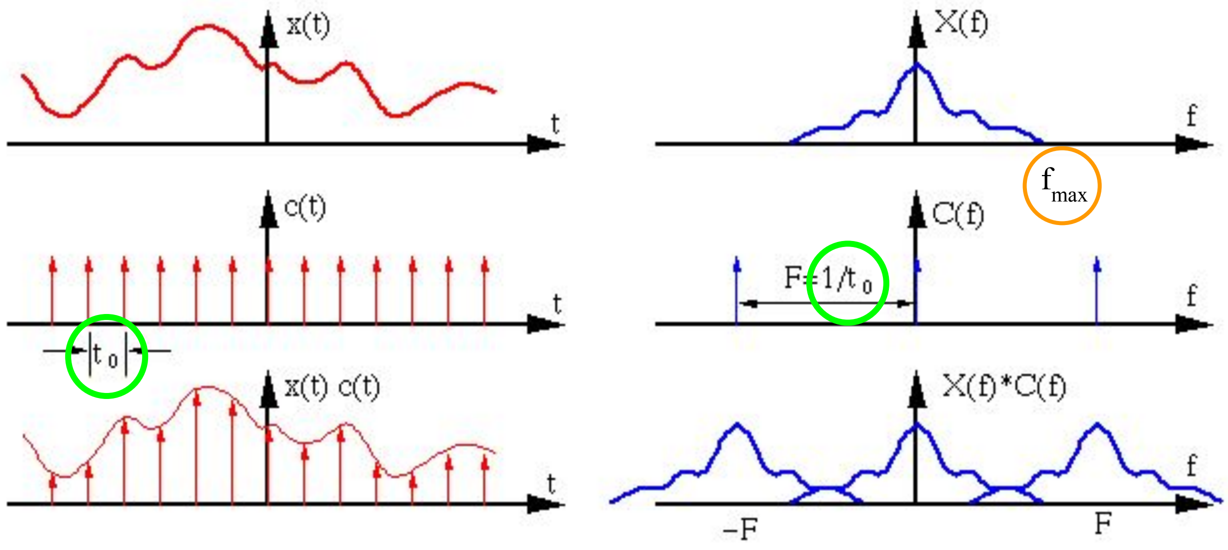


t_0 = pixel size

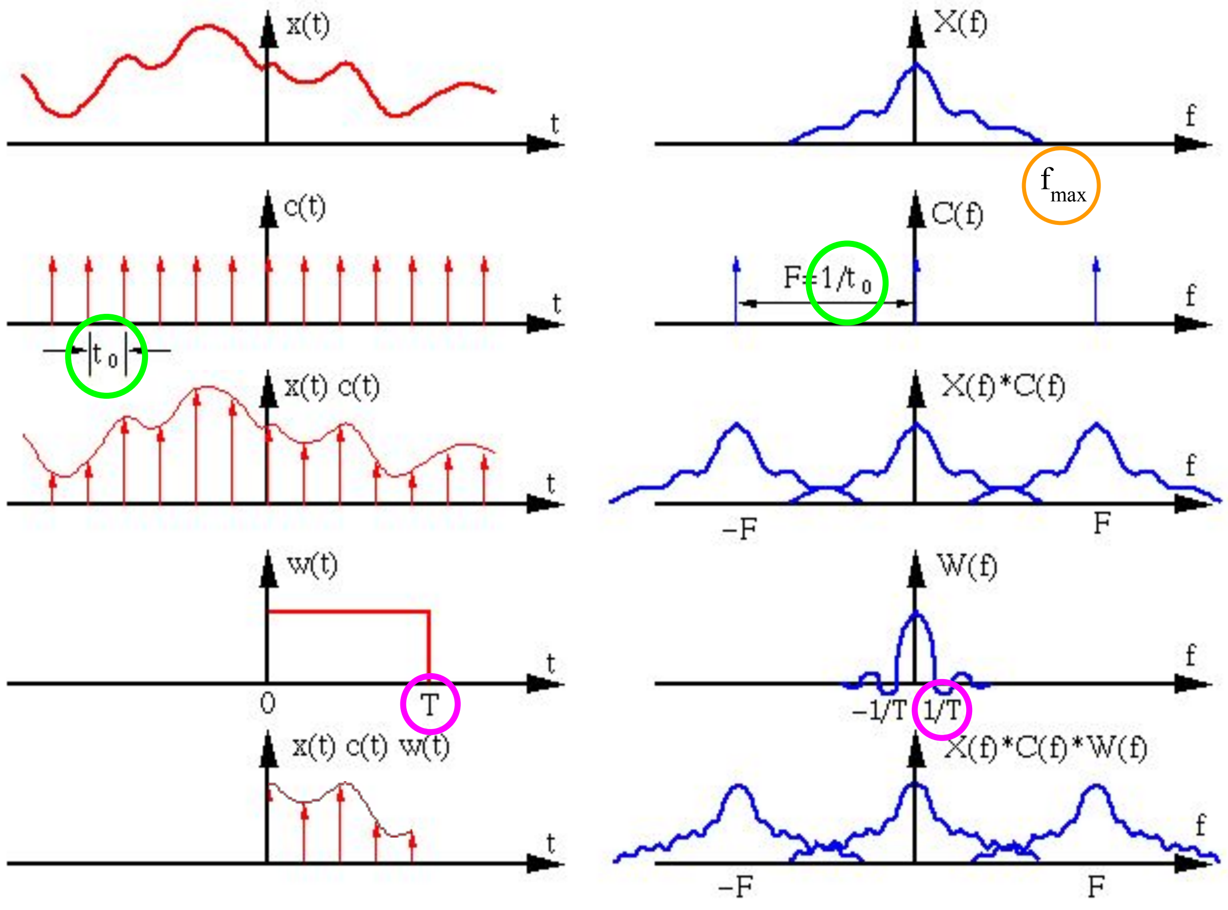
$1/t_0 >= 2\ f_{max}$ (to avoid frequency overlap)

$f_{max} <= 1/(2t_0)$ (Nyquist limit)

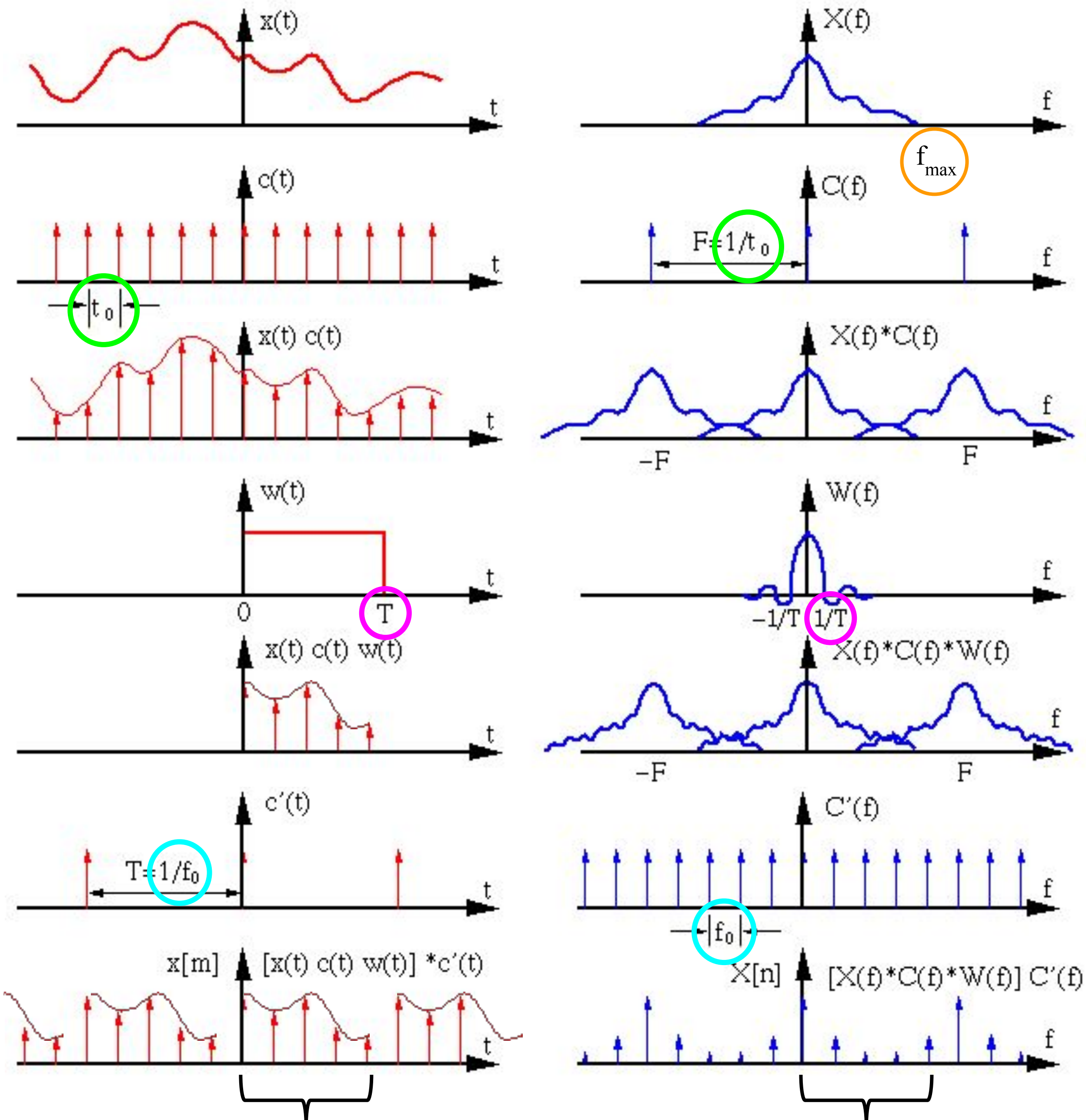
from: http://fourier.eng.hmc.edu/e101/lectures/Image_Processing/node2.html



from: http://fourier.eng.hmc.edu/e101/lectures/Image_Processing/node2.html



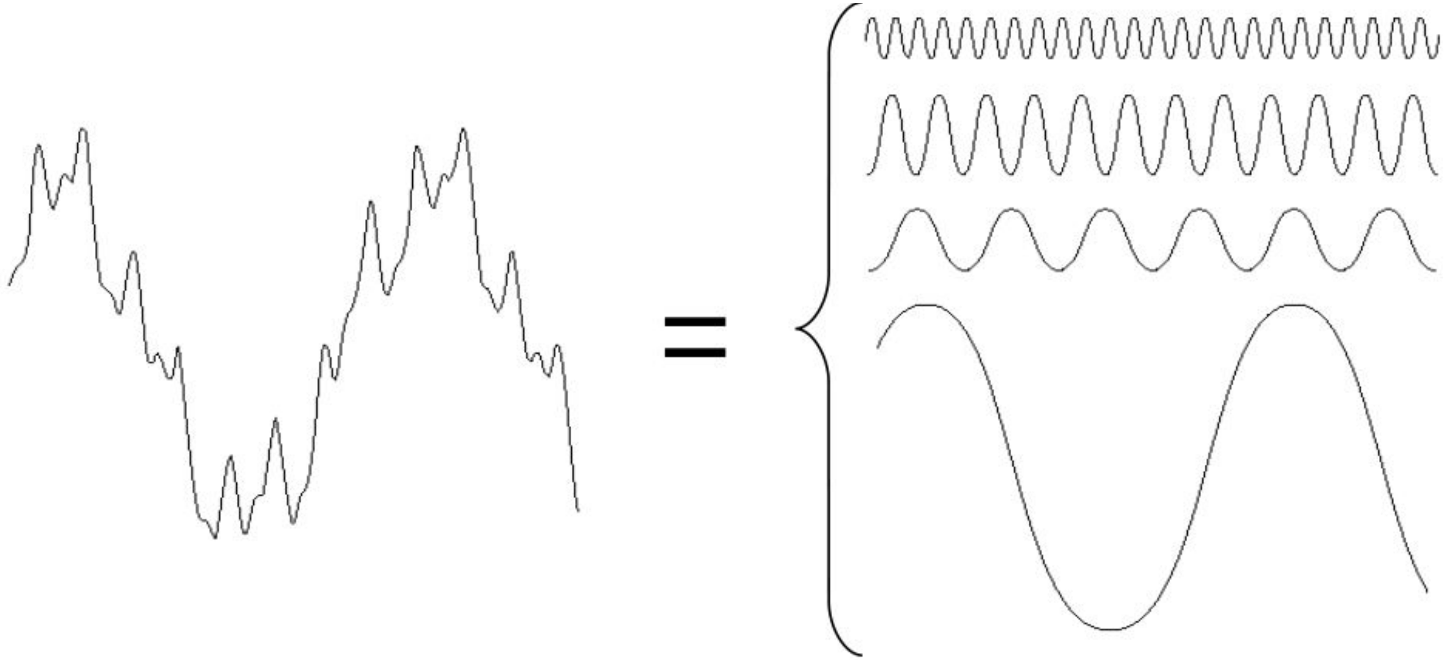
from: http://fourier.eng.hmc.edu/e101/lectures/Image_Processing/node2.html



Numerically sampled values

from: http://fourier.eng.hmc.edu/e101/lectures/Image_Processing/node2.html

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n cos(nf_0 t) + \sum_{n=1}^{\infty} b_n sin(nf_0 t)$$



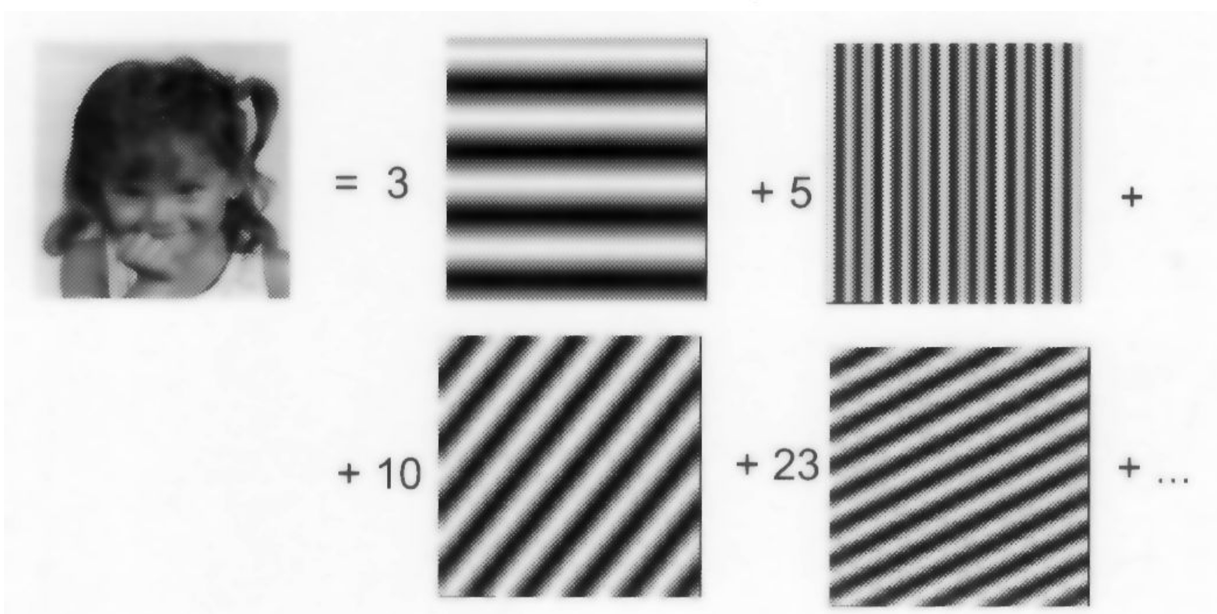
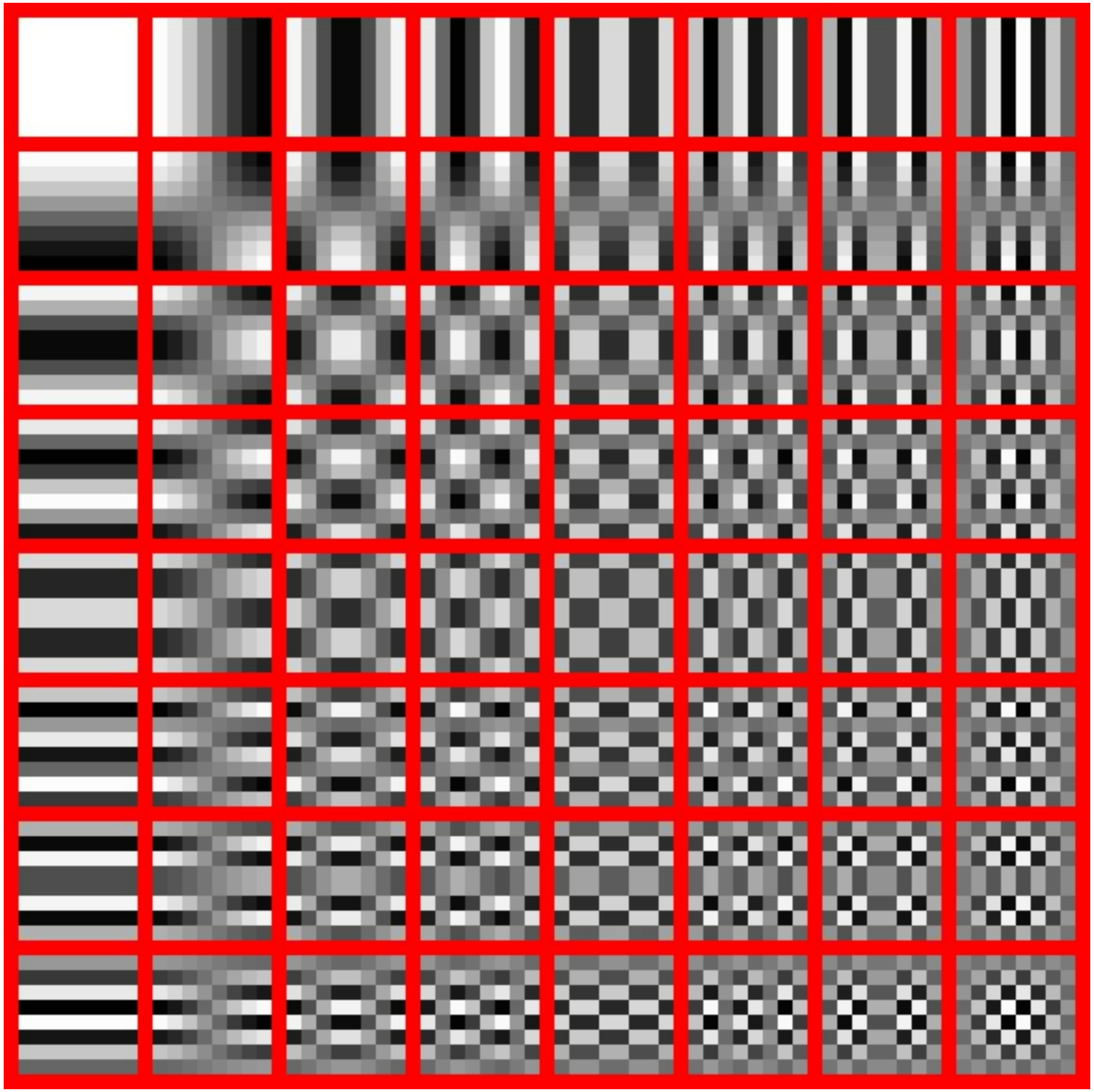
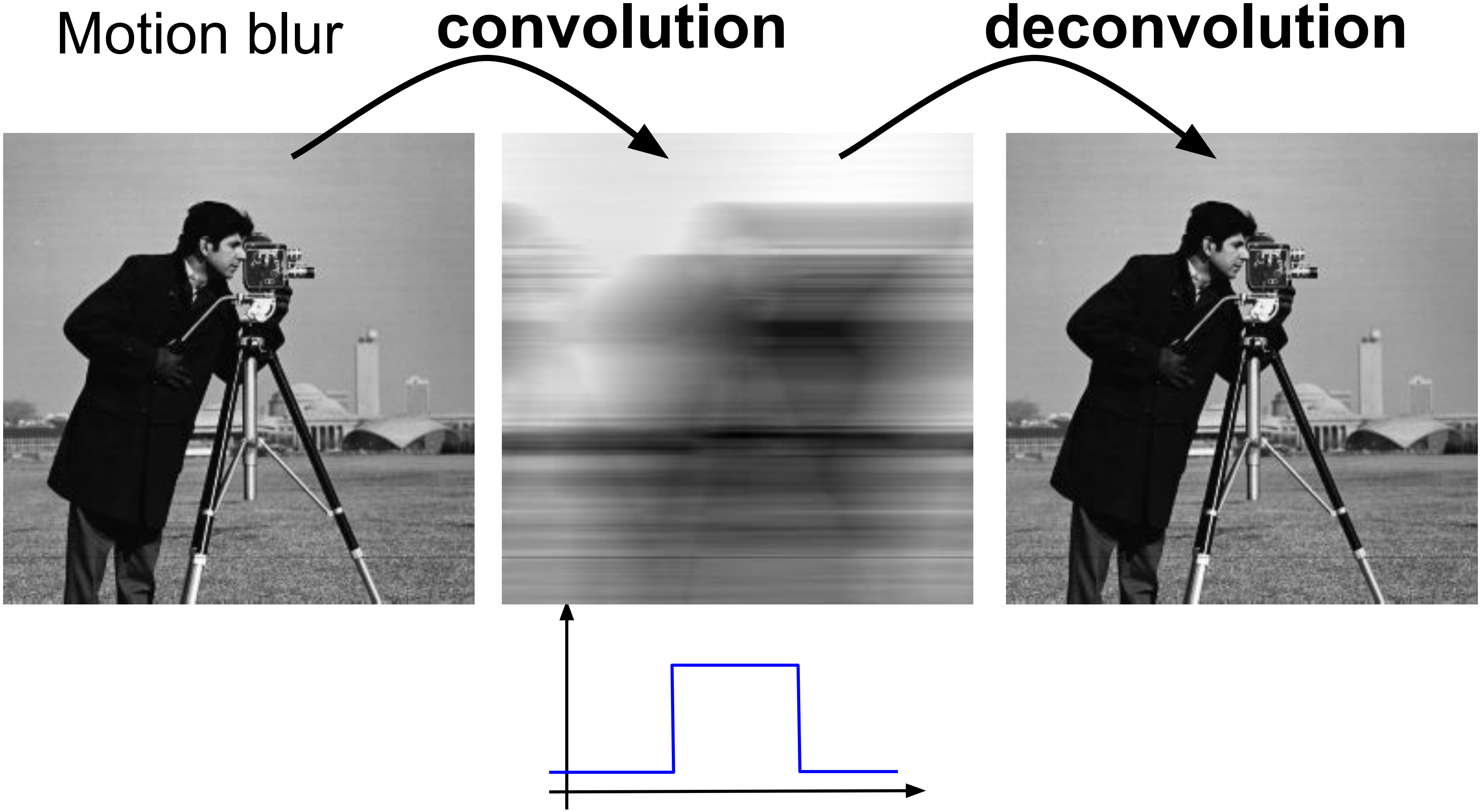


Image frequency decompostion

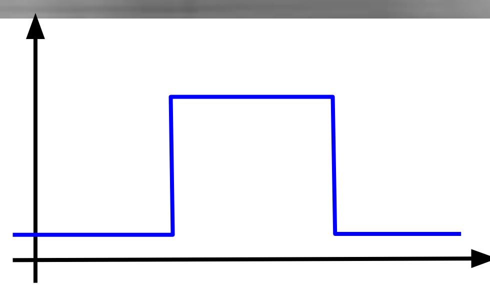


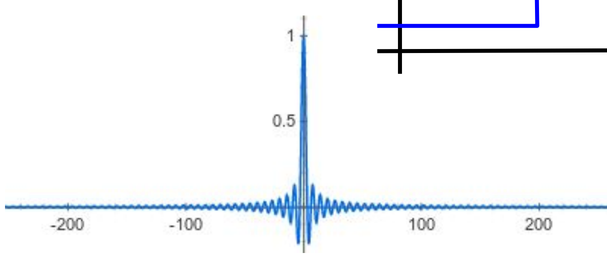
Resulting image	Weighted component	Fourier component

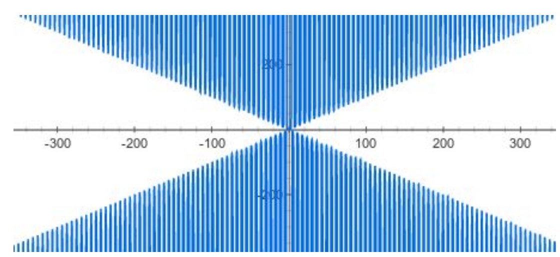
Motion blur


Motion blur
convolution
deconvolution

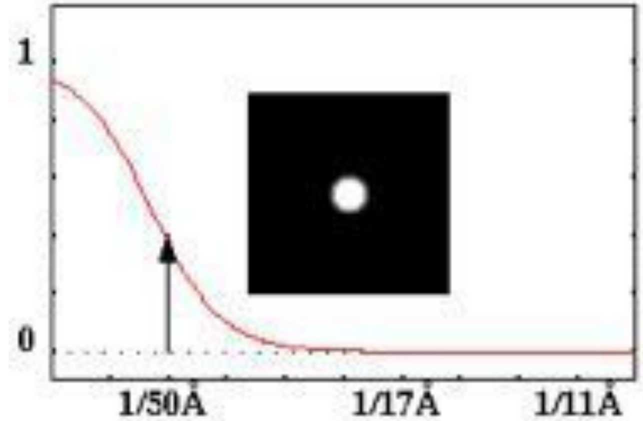
Motion blur **convolution** **deconvolution**





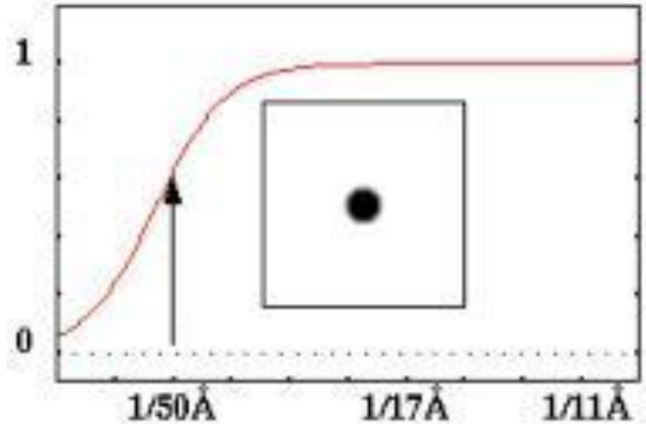


Fourier space filtering



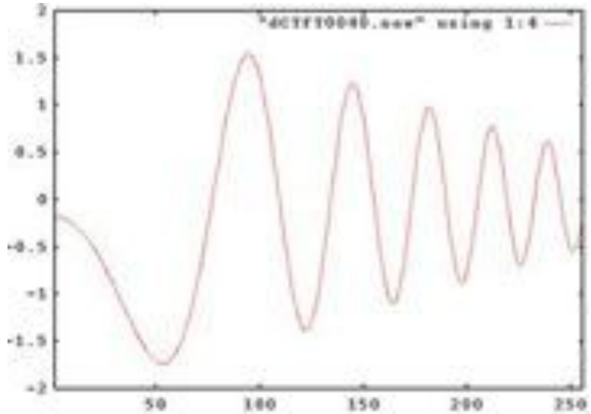
low-pass

Fourier space filtering



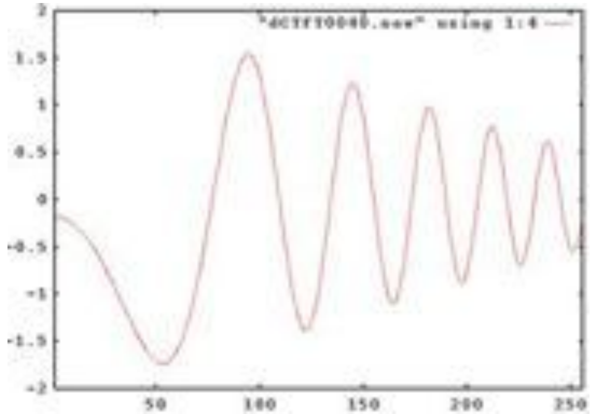
high-pass

Fourier space filtering

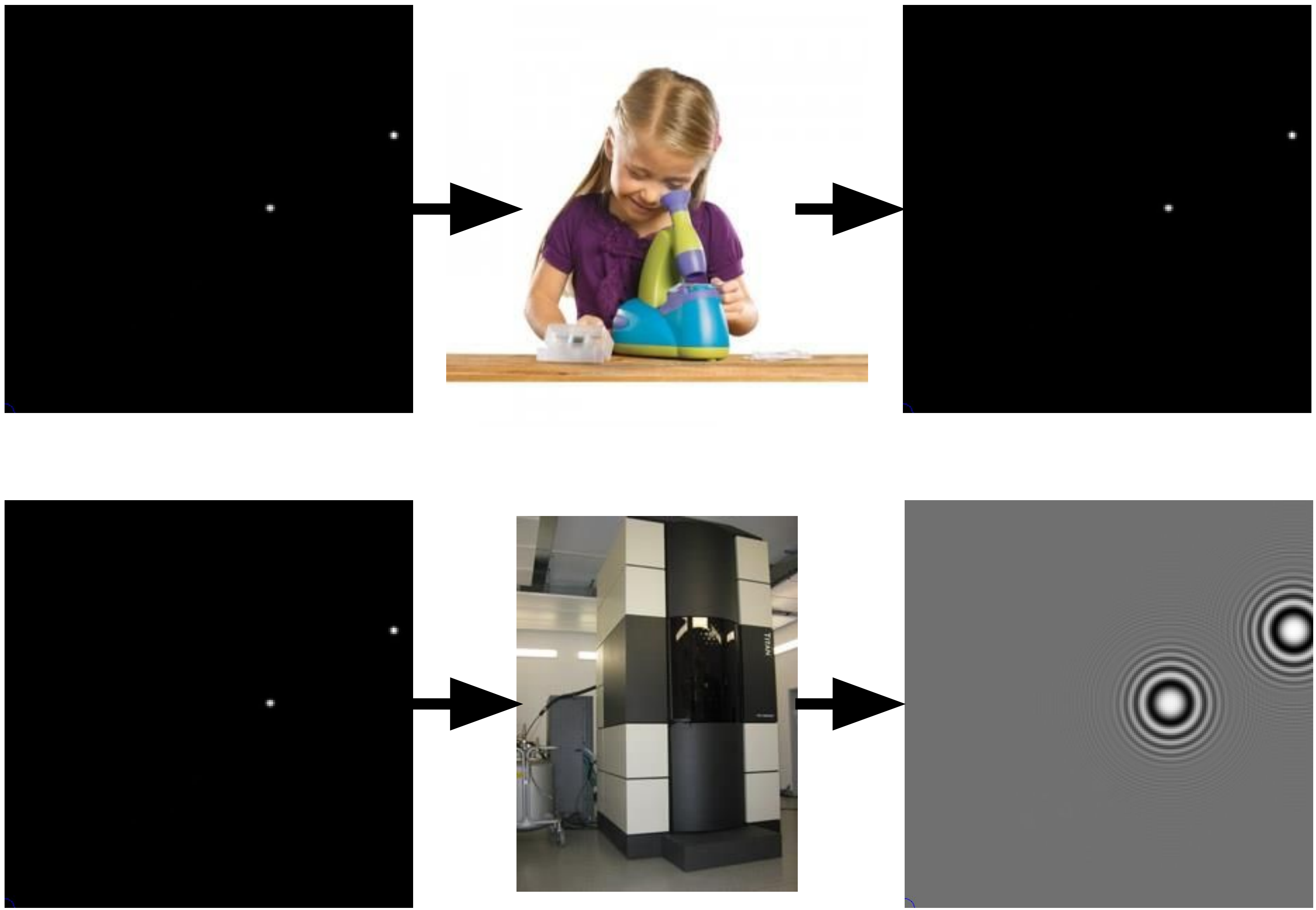


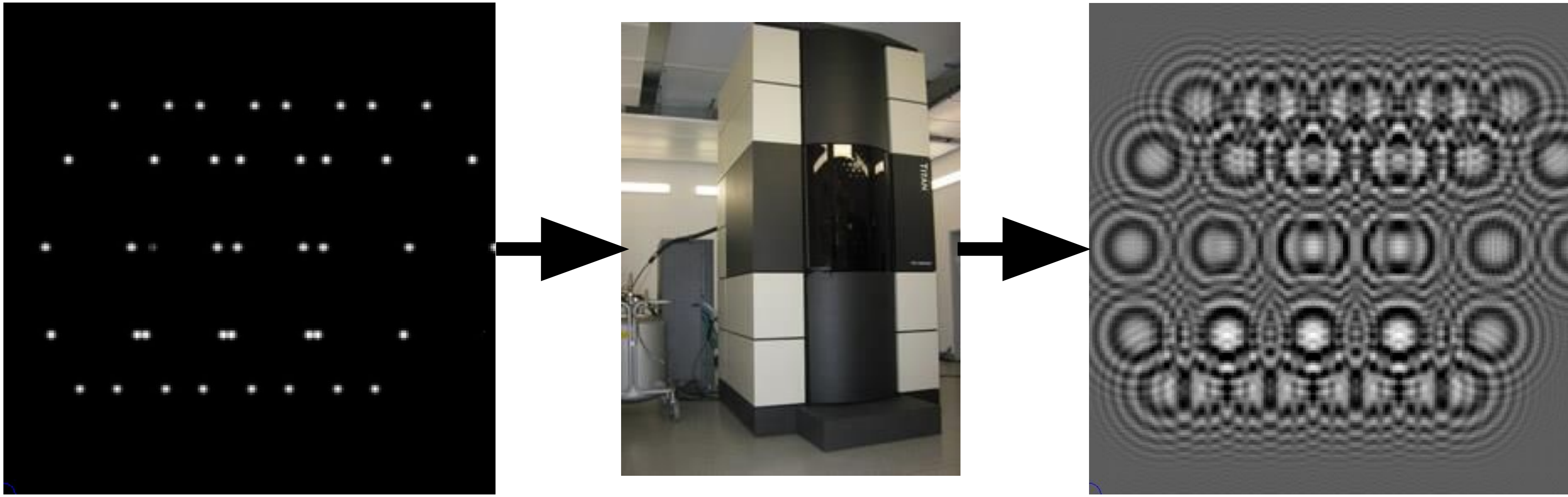
CTF

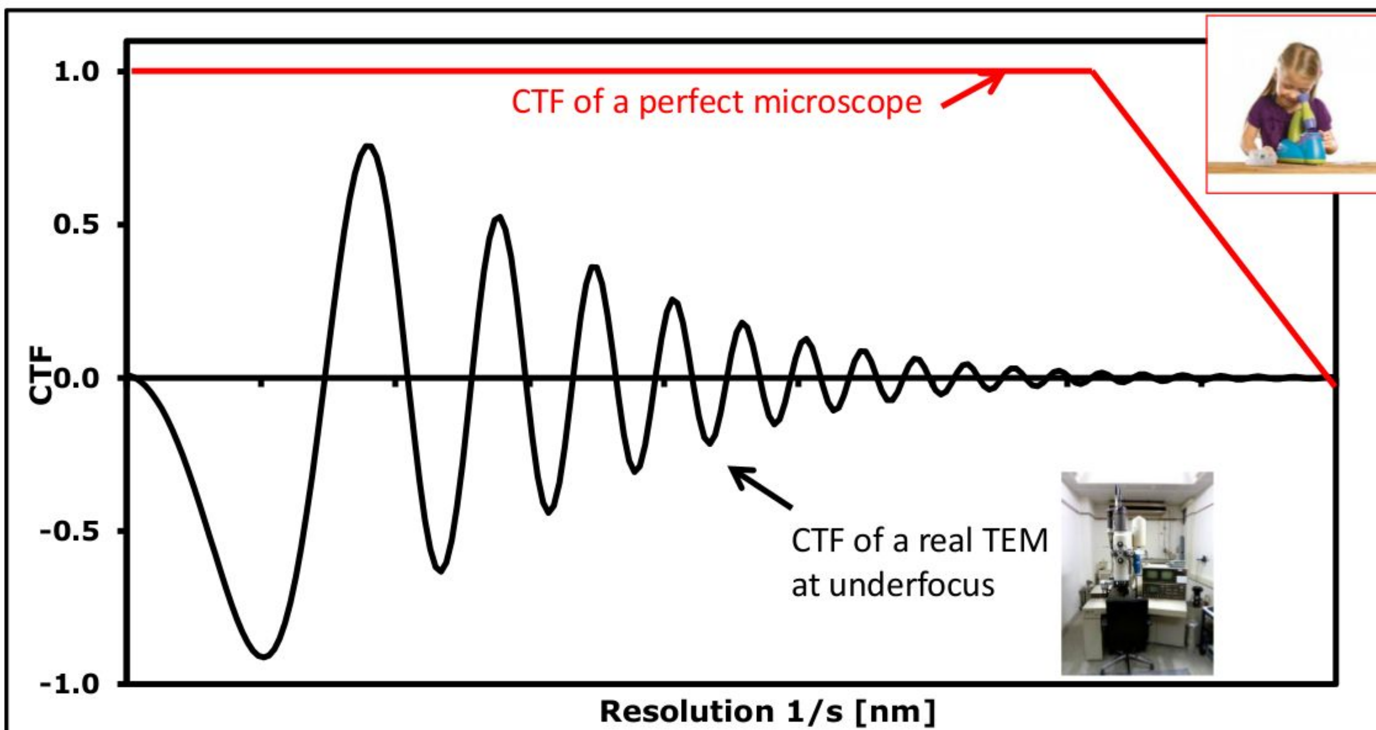
Fourier space filtering

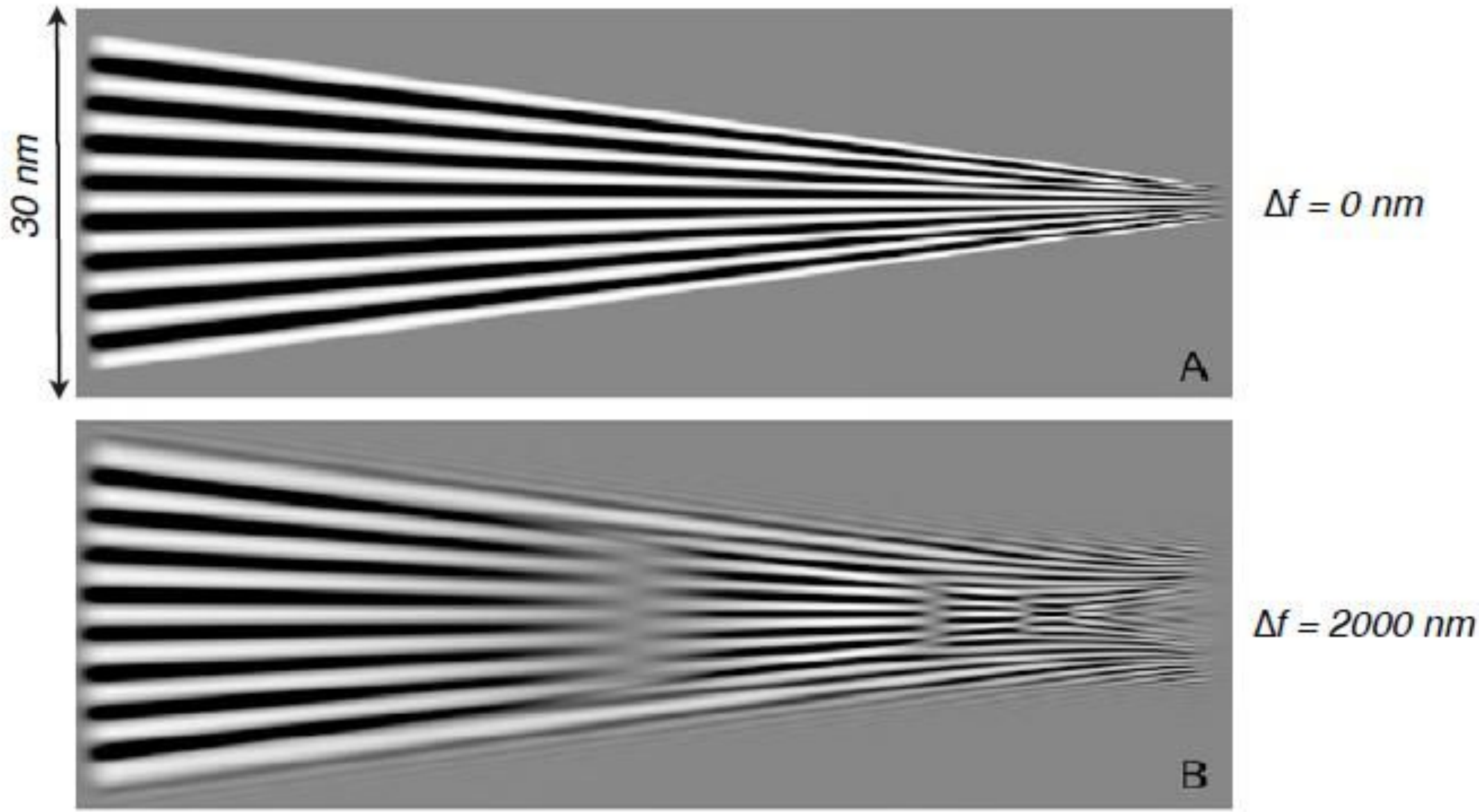


CTF-corrected


TITAN

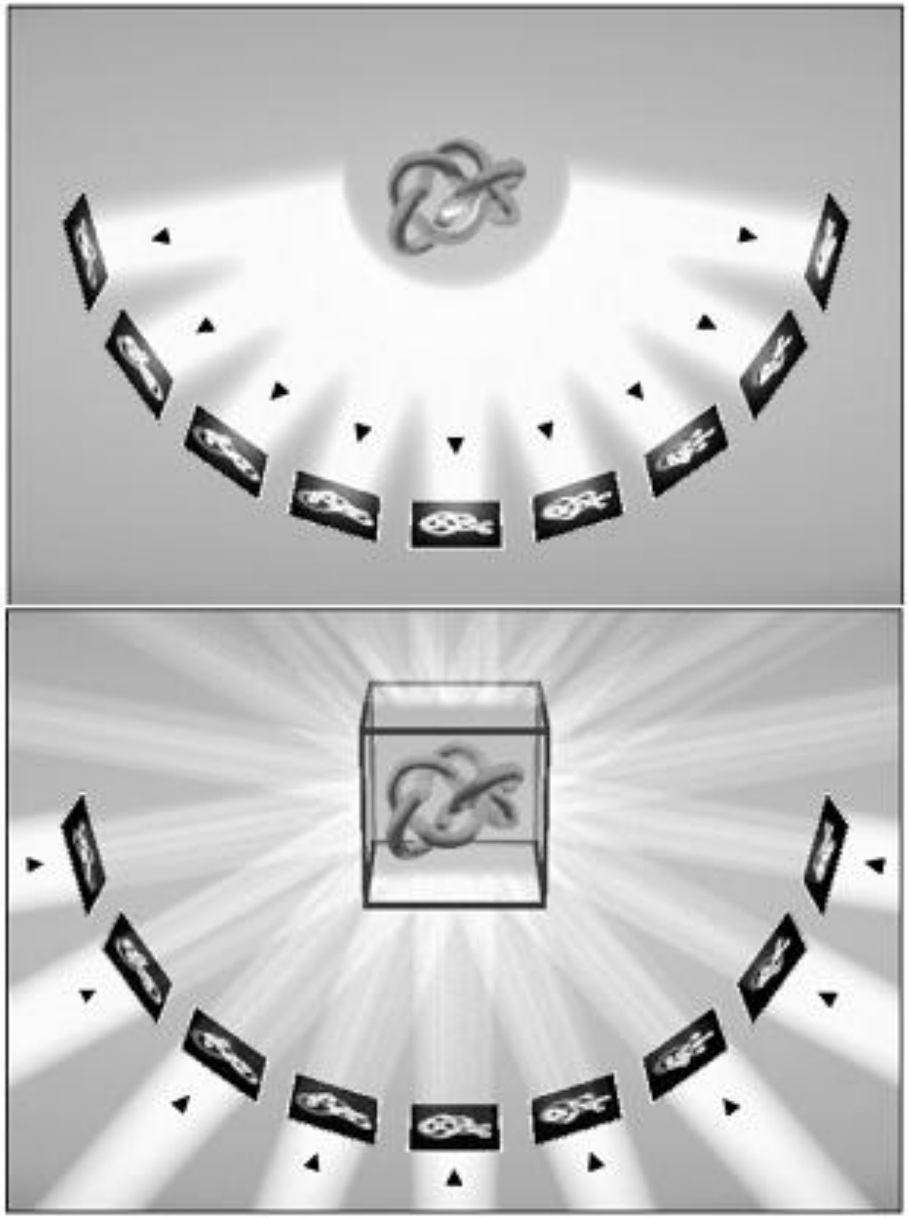



1.0
0.5
0.0
-0.5
-1.0
CTF
CTF of a perfect microscope
CTF of a real TEM
at underfocus
Resolution 1/s [nm]



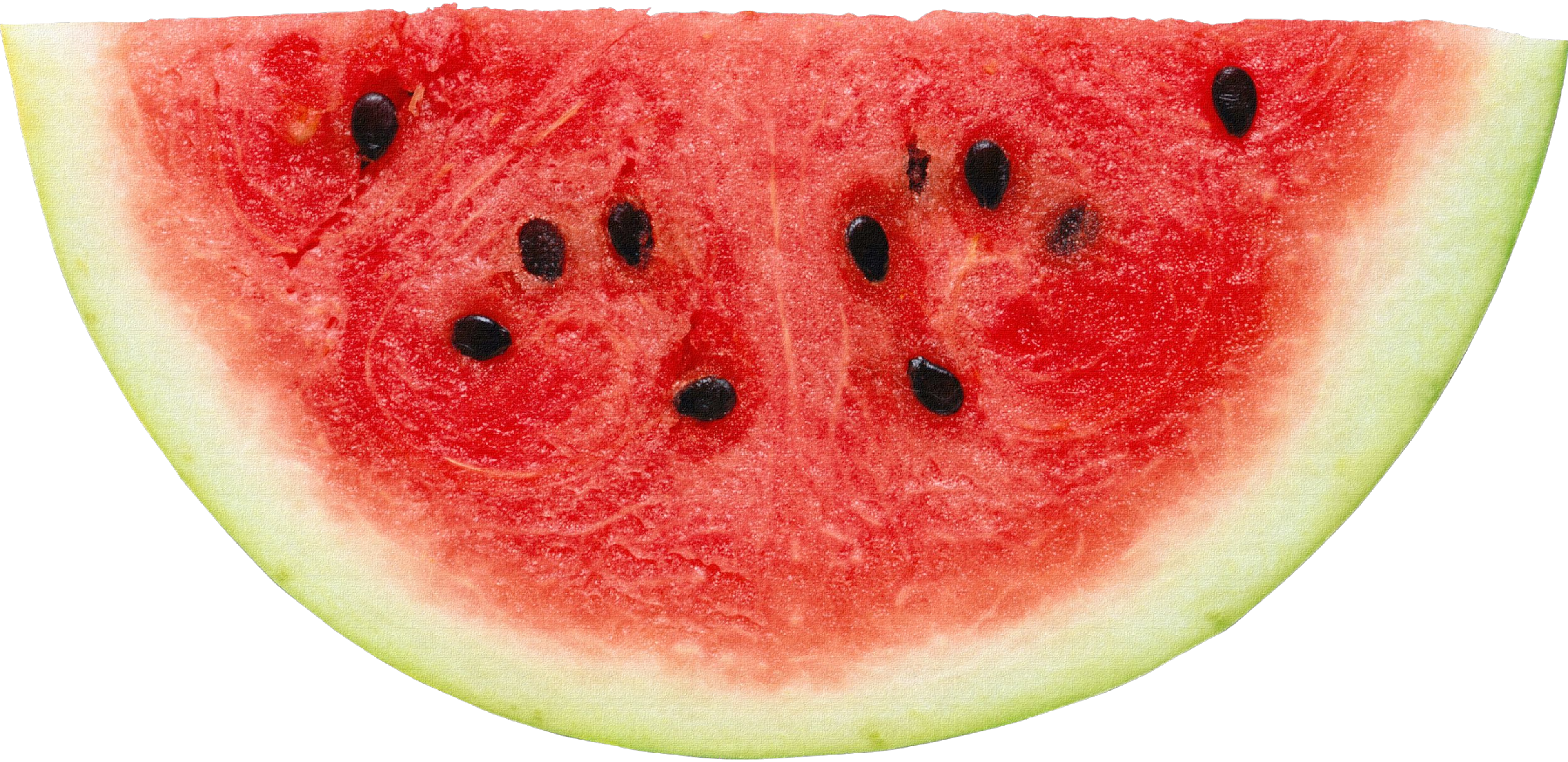
Downing & Glaeser, Ultramicroscopy 2008

Projections and retro-projections



Projection theorem (or Fourier slice)

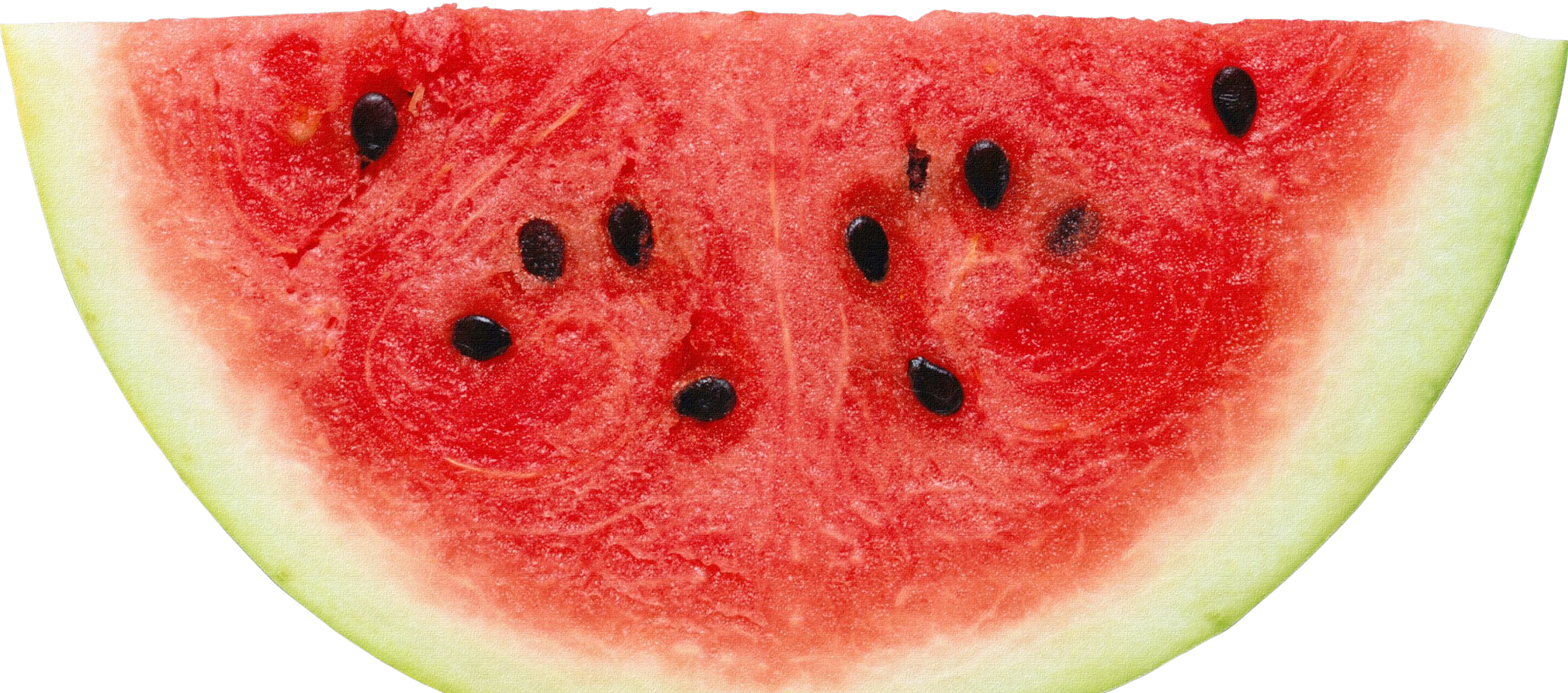
$$F(h,k,l) = \iiint_{-\infty}^{\infty} \rho(x,y,z) e^{2\pi i(xh+yk+zl)} dxdydz$$



Projection theorem (or Fourier slice)

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \rho(x,y,z) e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \overbrace{\int_{-\infty}^{\infty} \rho(x,y,z) dz}\, e^{2\pi i(xh+yk+zl)} dxdydz$$



Projection theorem (or Fourier slice)

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \rho(x,y,z) e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \overbrace{\underbrace{\int_{-\infty}^{\infty} \rho(x,y,z)dz}} \; e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iiint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk+zl)} dxdydz$$

Projection theorem (or Fourier slice)

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \rho(x,y,z) e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x,y,z) dz \; e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iiint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk)} dxdy \int_{-\infty}^{\infty} e^{2\pi i(zl)} dz$$

Projection theorem (or Fourier slice)

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \rho(x,y,z) e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \overbrace{\underbrace{\int_{-\infty}^{\infty} \rho(x,y,z)dz}}\ e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iiint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk)} dxdy \underbrace{\int_{-\infty}^{\infty} e^{2\pi i(zl)} dz}$$

$$F(h,k,l) = \iint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk)} dxdy\ \delta(l)$$

Projection theorem (or Fourier slice)

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \rho(x,y,z) e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iiint_{-\infty}^{\infty} \overbrace{\underbrace{\int_{-\infty}^{\infty} \rho(x,y,z)dz}}\; e^{2\pi i(xh+yk+zl)} dxdydz$$

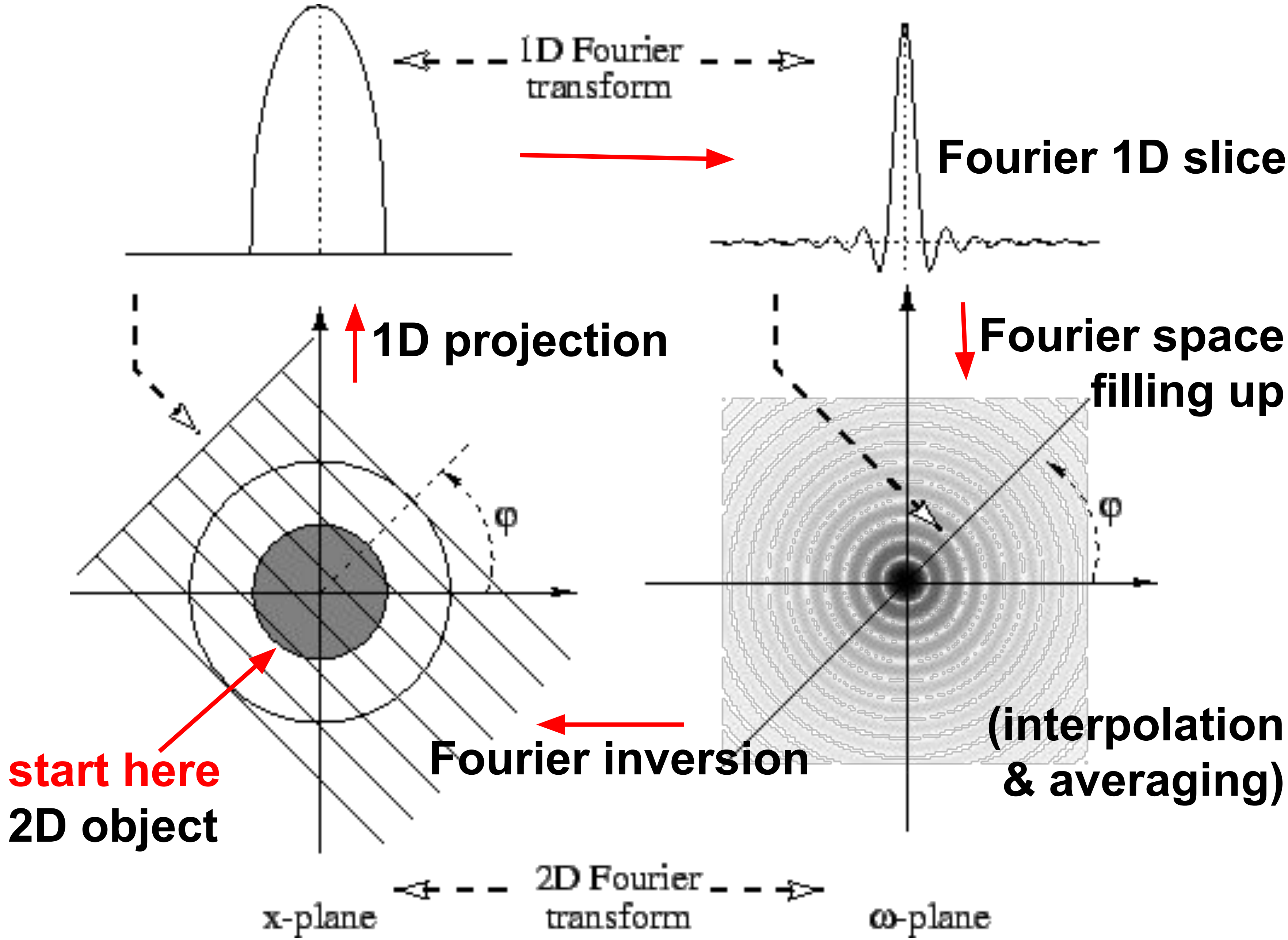
$$F(h,k,l) = \iiint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk+zl)} dxdydz$$

$$F(h,k,l) = \iint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk)} dxdy \underbrace{\int_{-\infty}^{\infty} e^{2\pi i(zl)} dz}$$

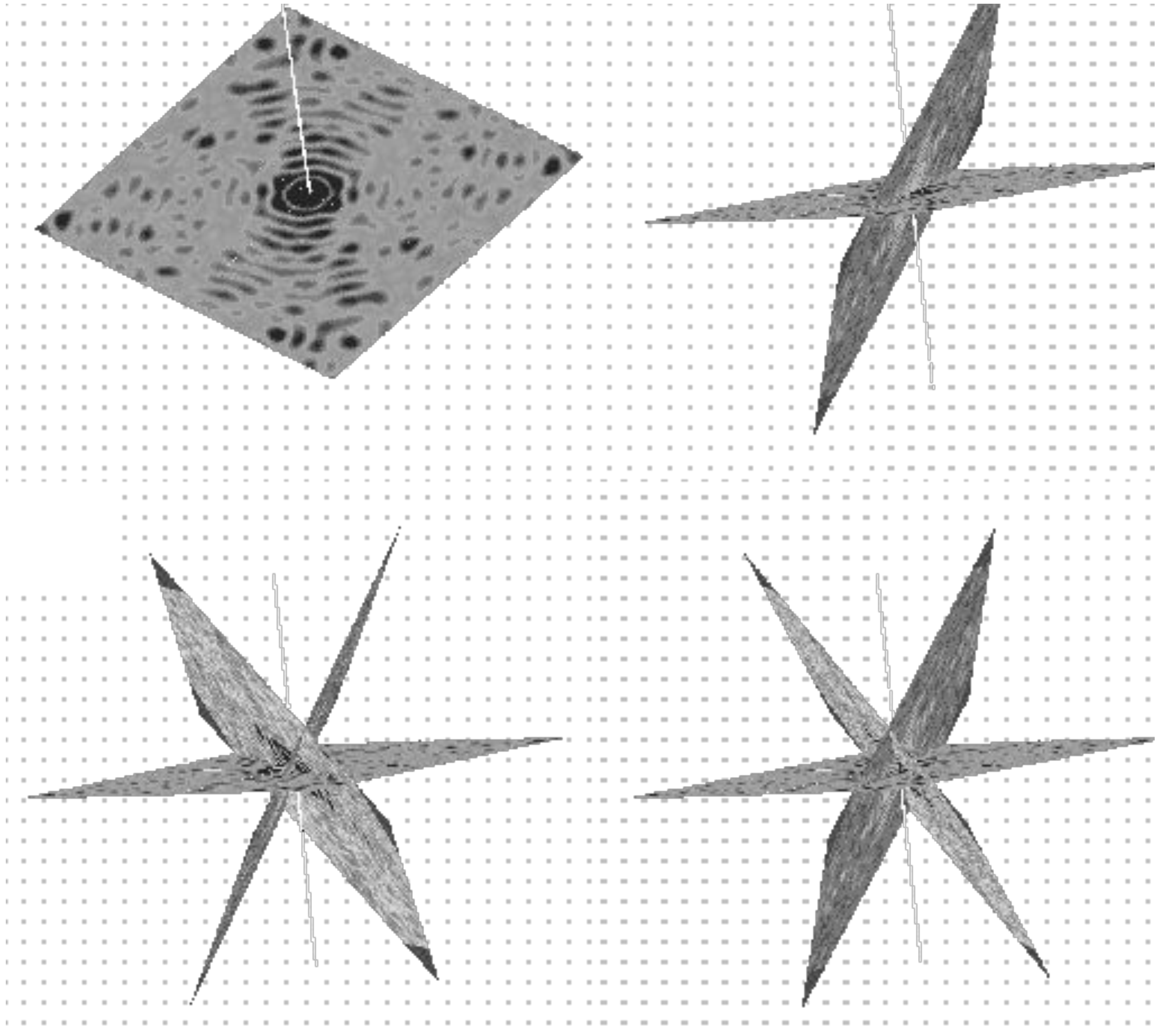
$$F(h,k,l) = \iint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk)} dxdy\; \delta(l)$$

$$F(h,k,0) = \iint_{-\infty}^{\infty} Img(x,y) e^{2\pi i(xh+yk)} dxdy$$

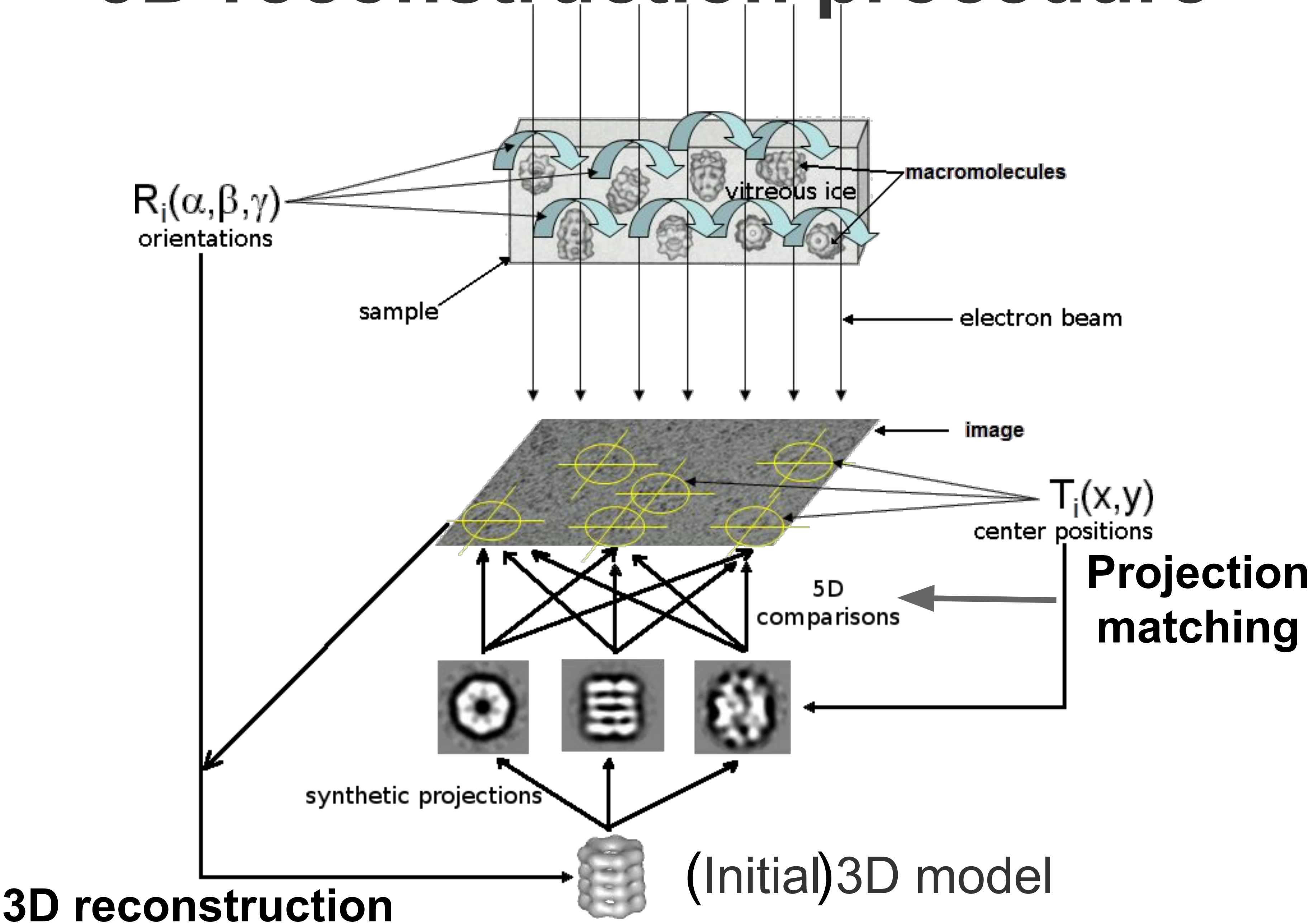
Projection (or Fourier slice) theorem 2D



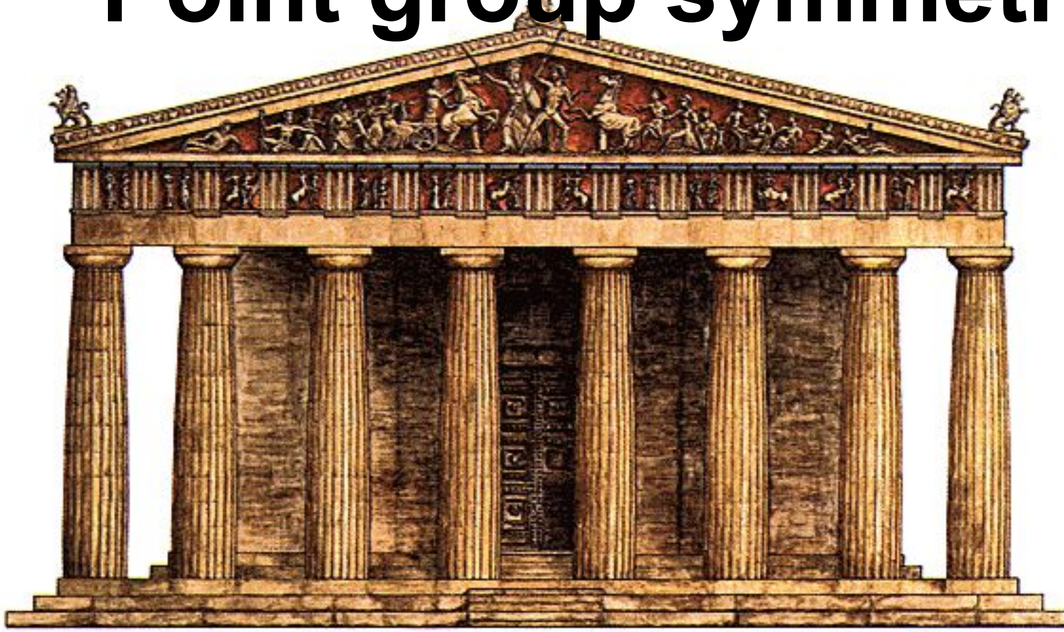
Projection (or Fourier slice) theorem 3D



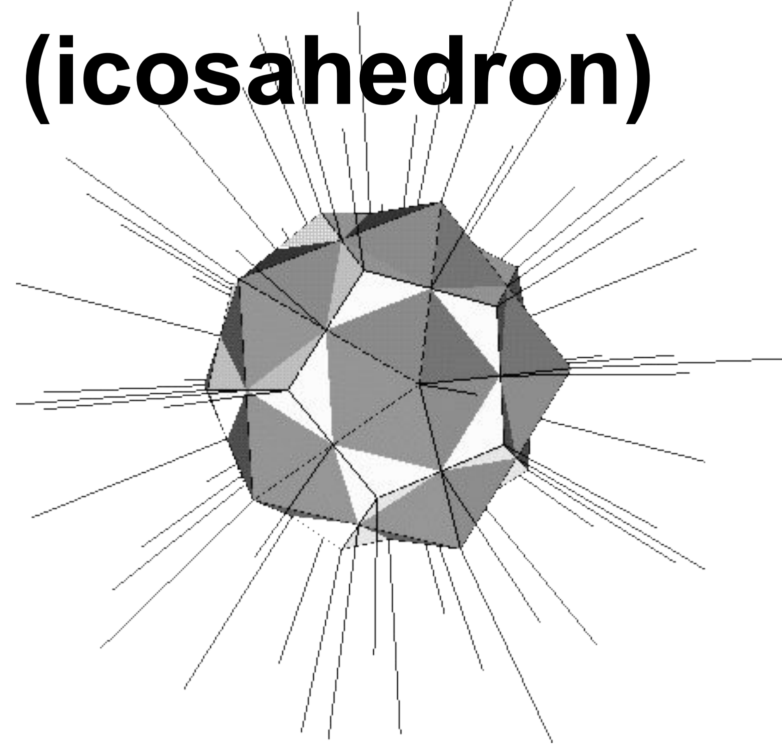
3D reconstruction procedure

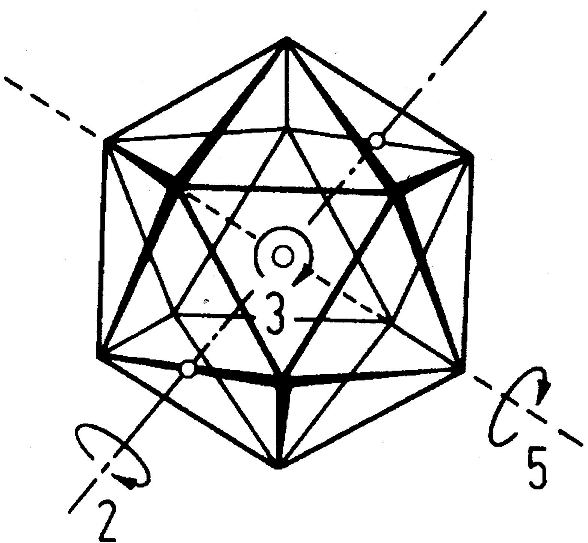


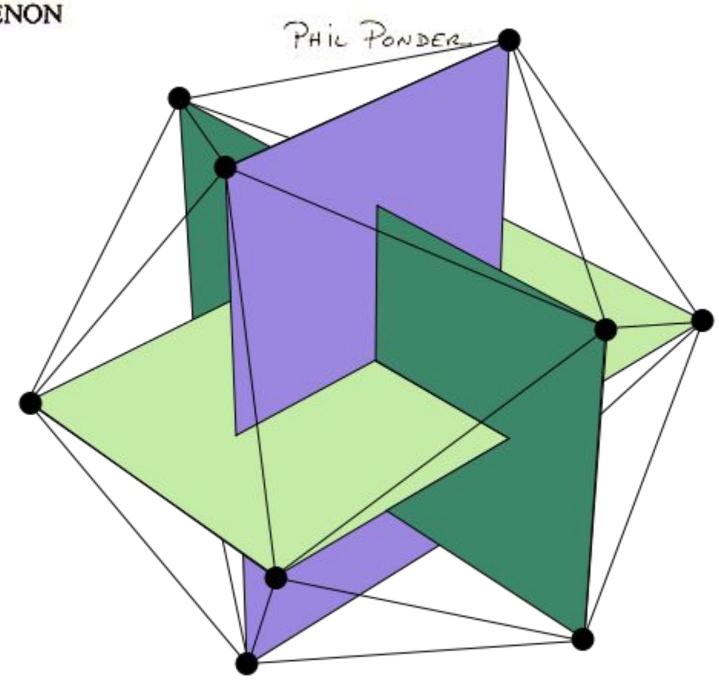
Point group symmetry (icosahedron)

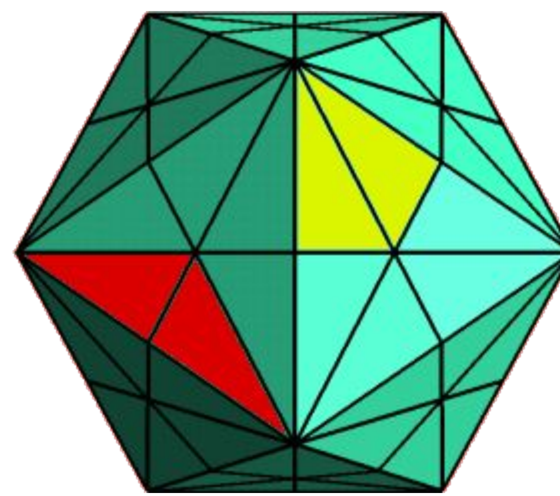


THE PARTHENON

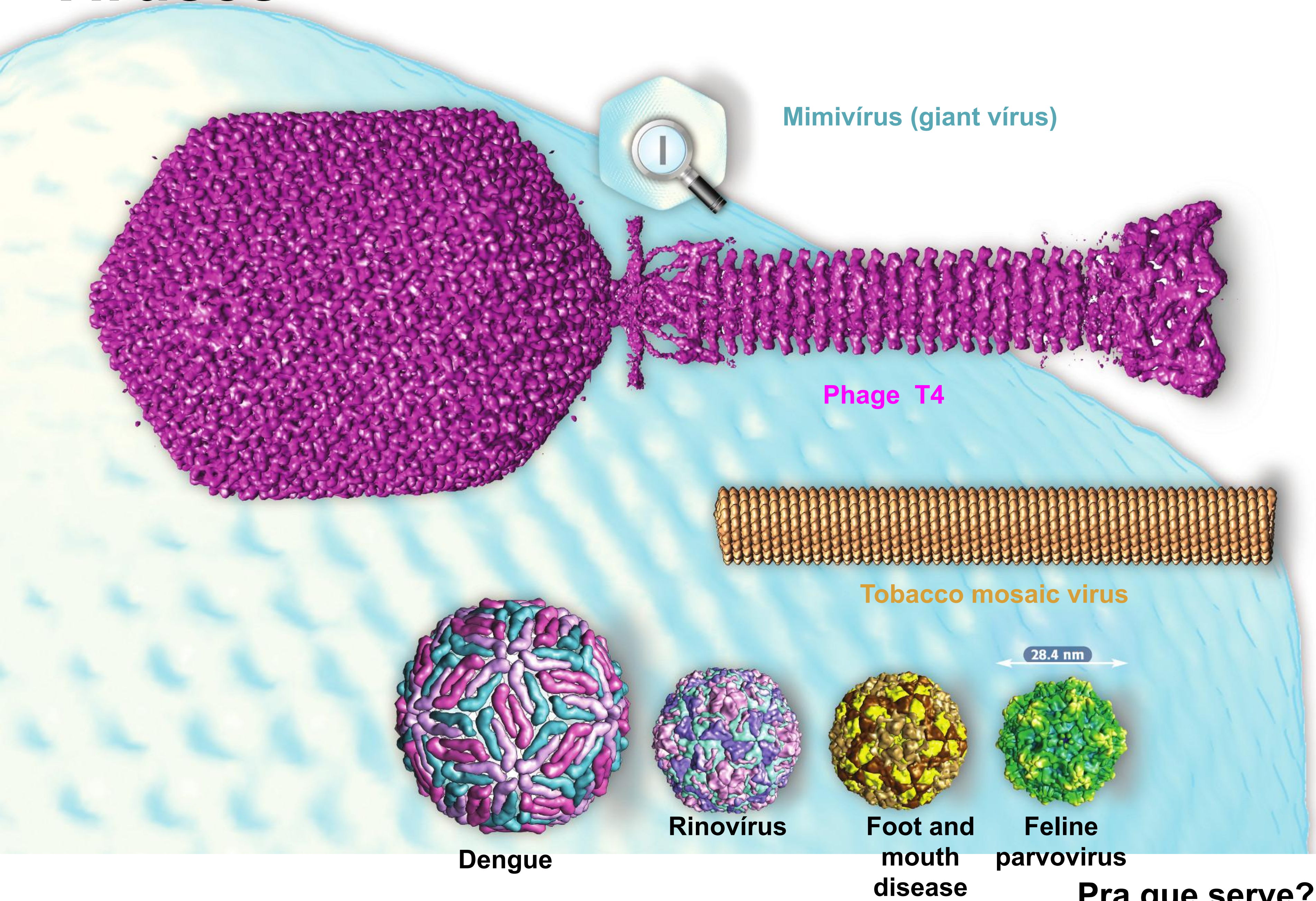




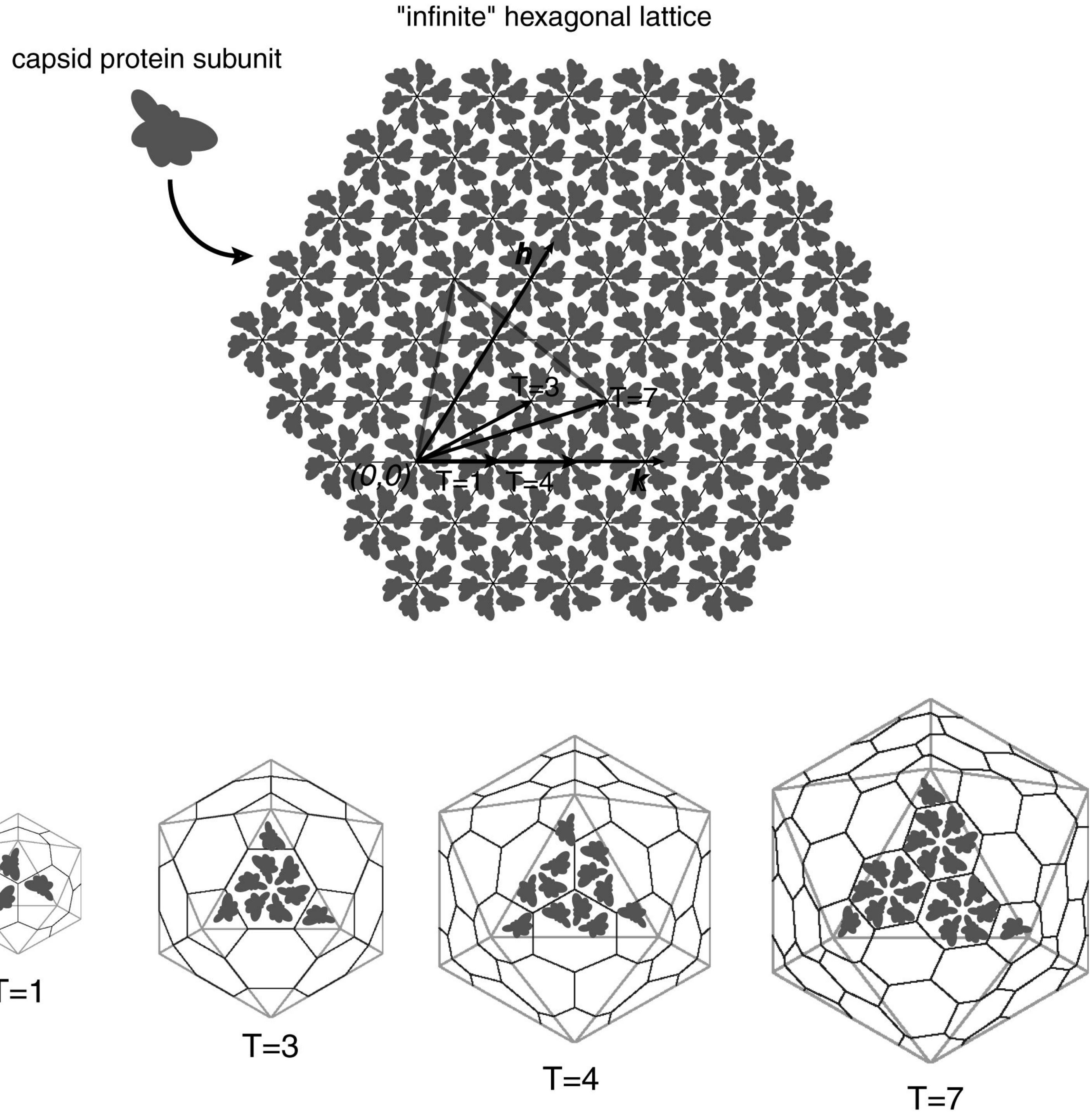


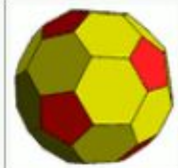
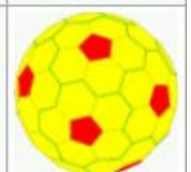
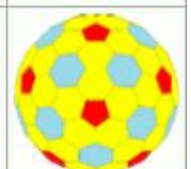
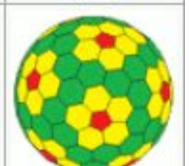
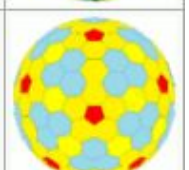


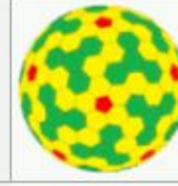
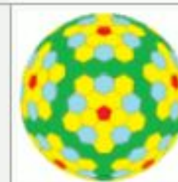
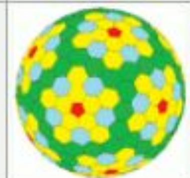
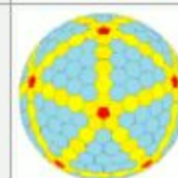
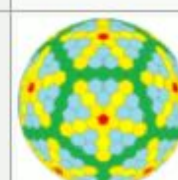
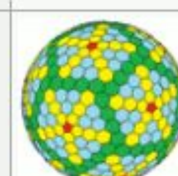
Viruses

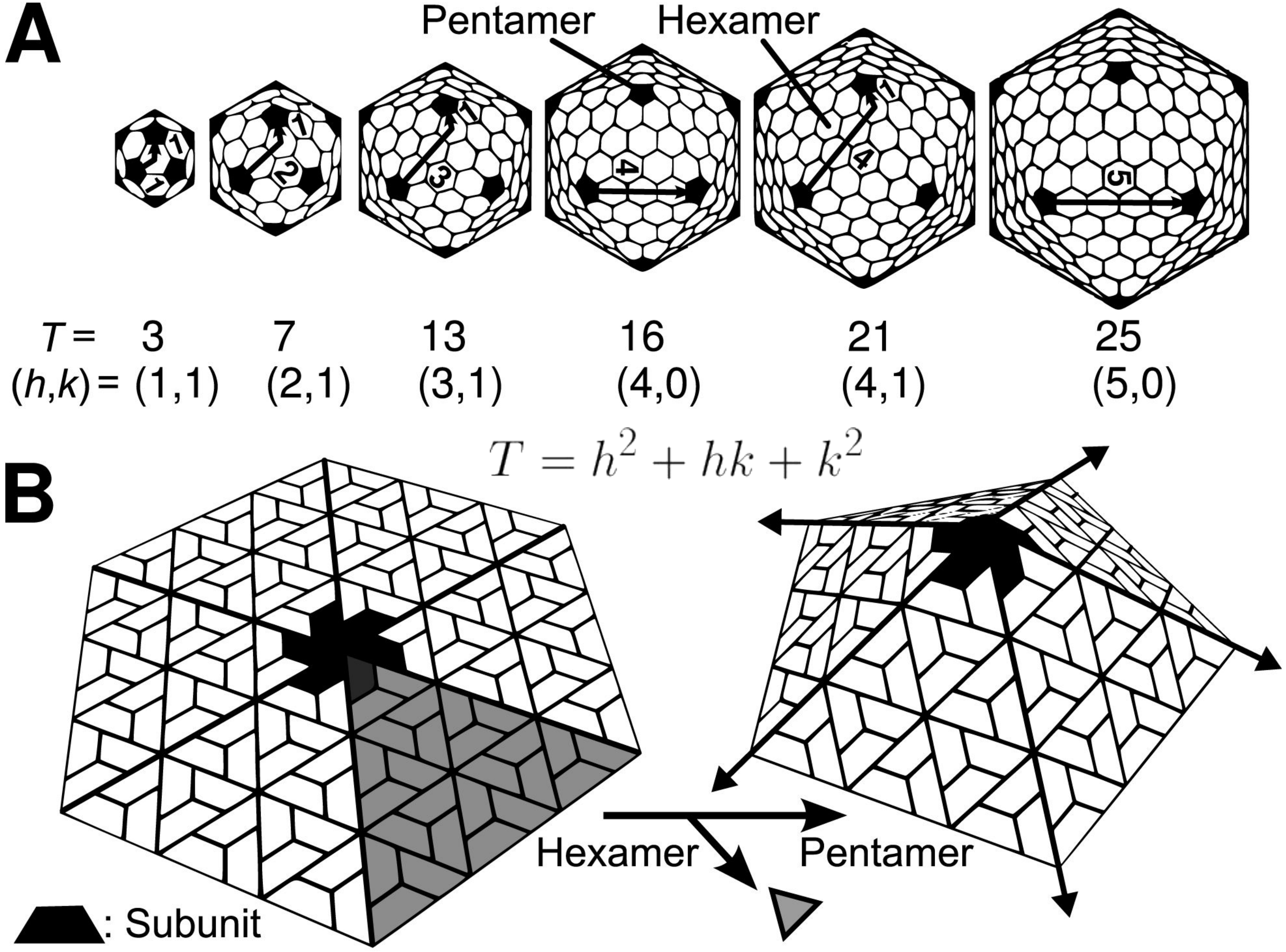


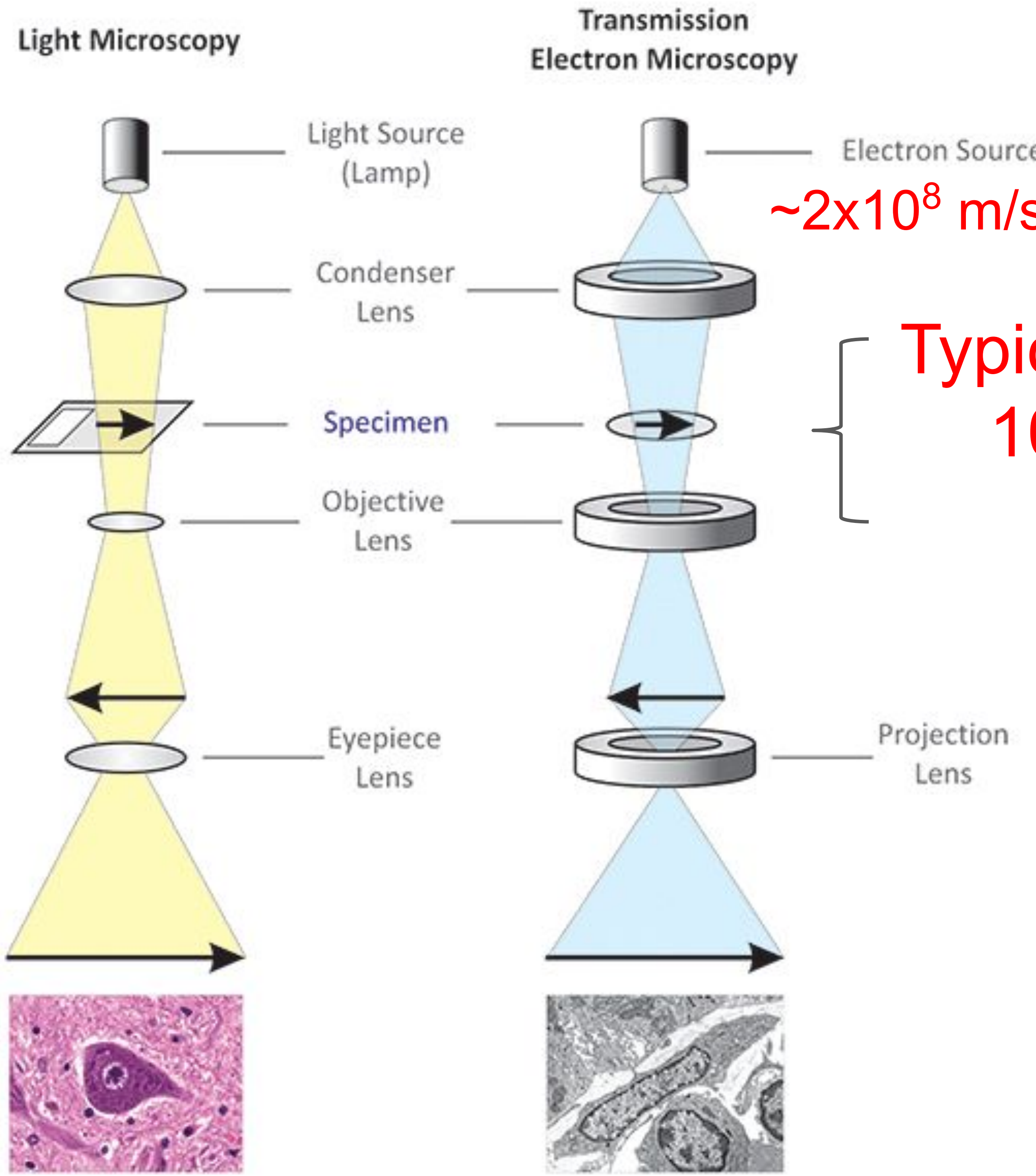
Pra que serve?


"infinite" hexagonal lattice
capsid protein subunit
h
T=3
T=7
(0,0)
T=1
T=4
k
T=1
T=3
T=4
T=7

(1,0)	1	
(1,1)	3	
(2,0)	4	
(2,1)	7	
(3,0)	9	
(2,2)	12	
(3,1)	13	
(4,0)	16	

(4,1)	
(5,0)	
(3,3)	
(4,2)	
(5,1)	
(6,0)	
(4,3)	
(5,2)	
(6,1)	
(4,4)	
(6,2)	
(5,3)	


A
Pentamer
Hexamer
T = 3 7 13 16 21 25
(h,k) = (1,1) (2,1) (3,1) (4,0) (4,1) (5,0)
$T = h^2 + hk + k^2$
B
Hexamer
Pentamer
: Subunit



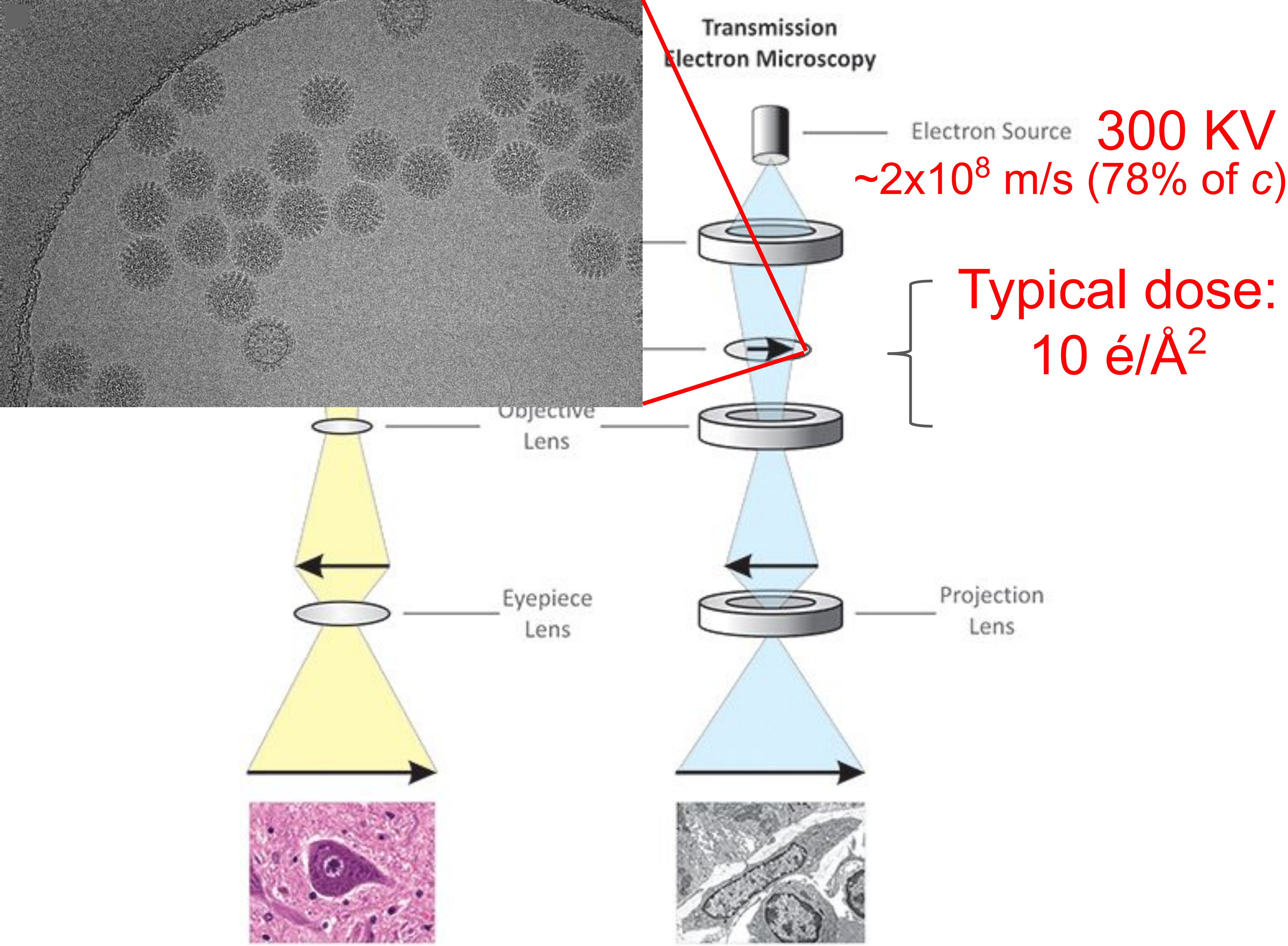
300 KV

~2×10^8 m/s (78% of *c*)

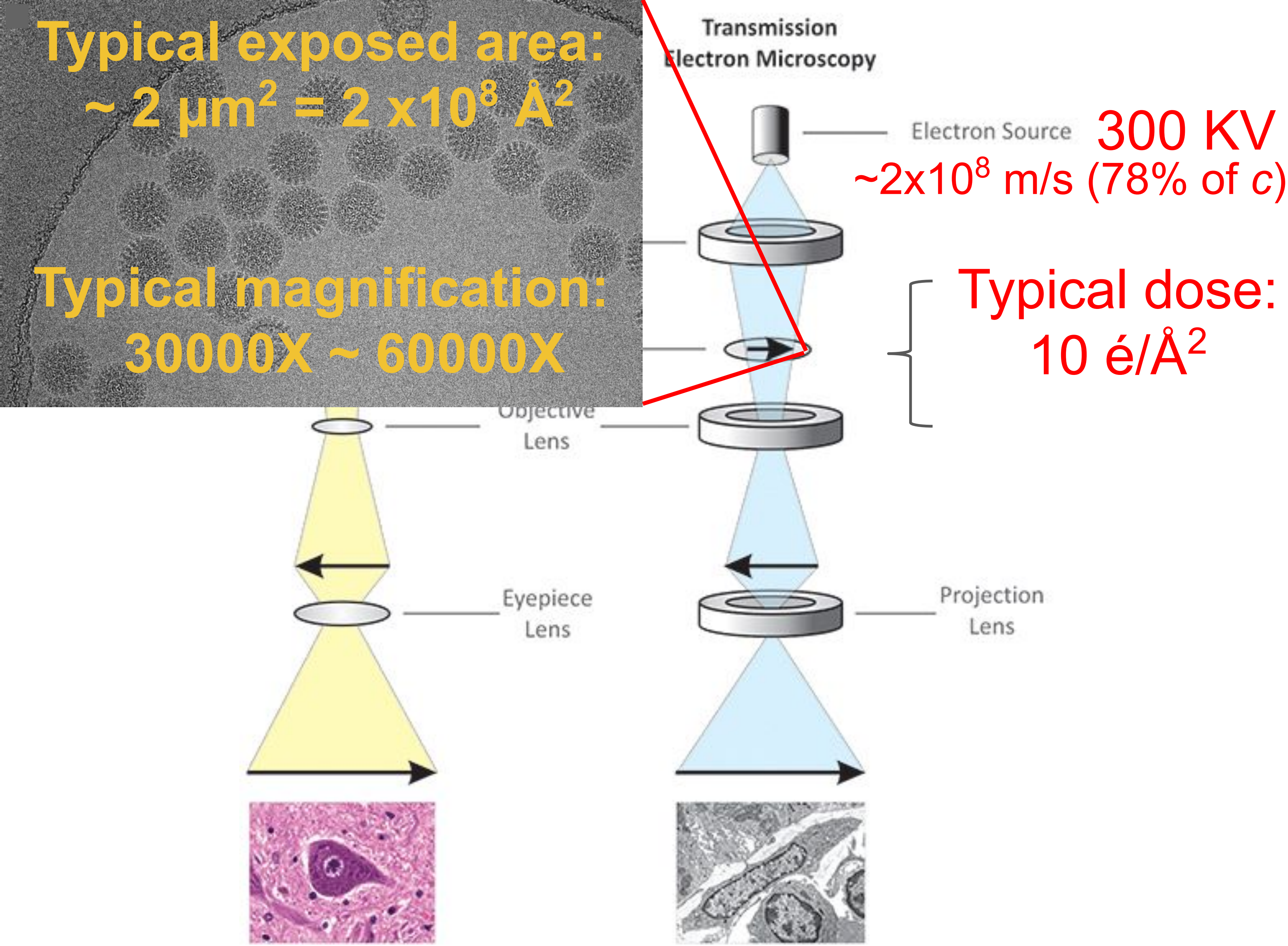
Typical dose:

10 é/Å^2

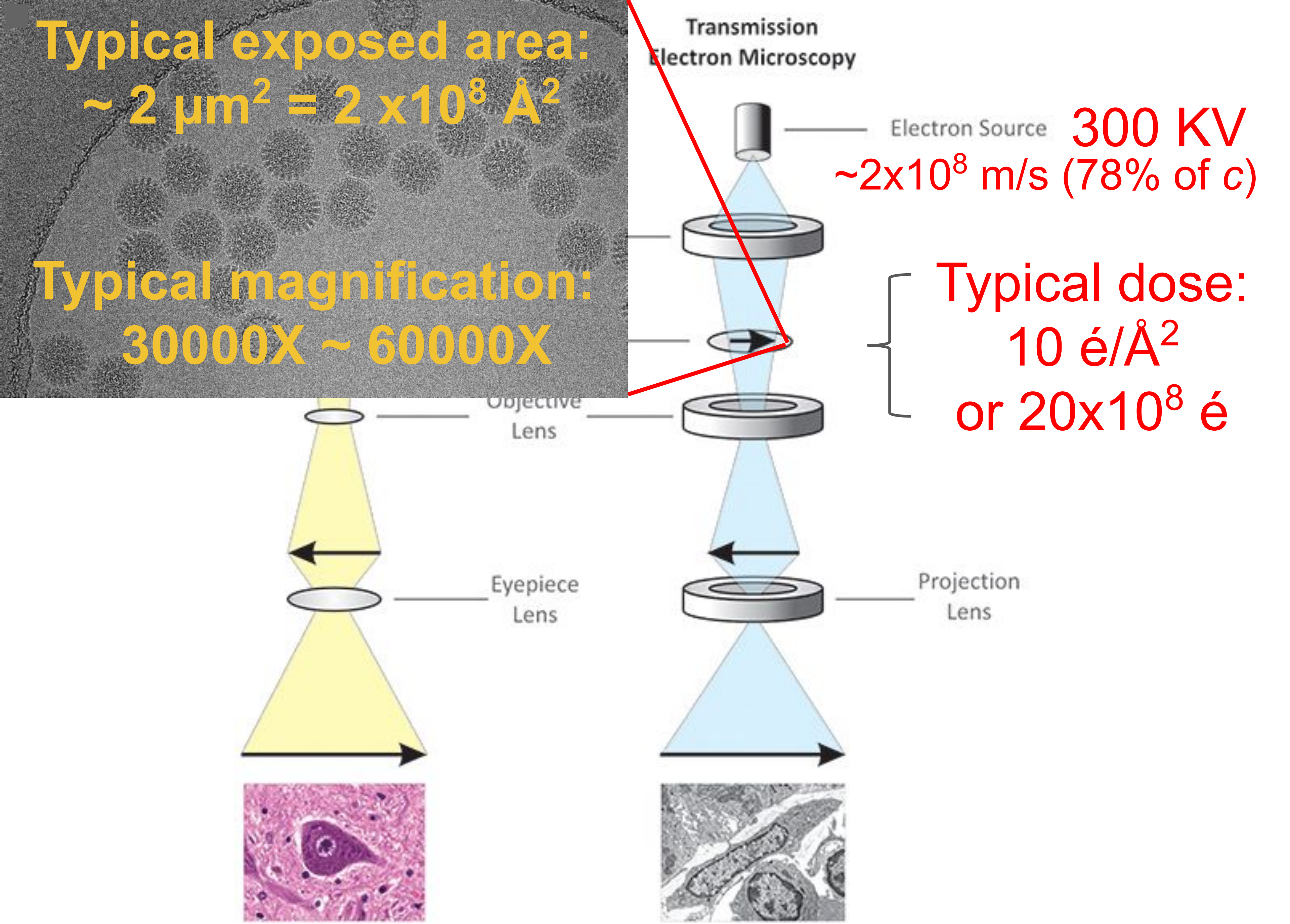
From: http://www.microbiologyinfo.com/differences-between-light-microscope-and-electron-microscope/



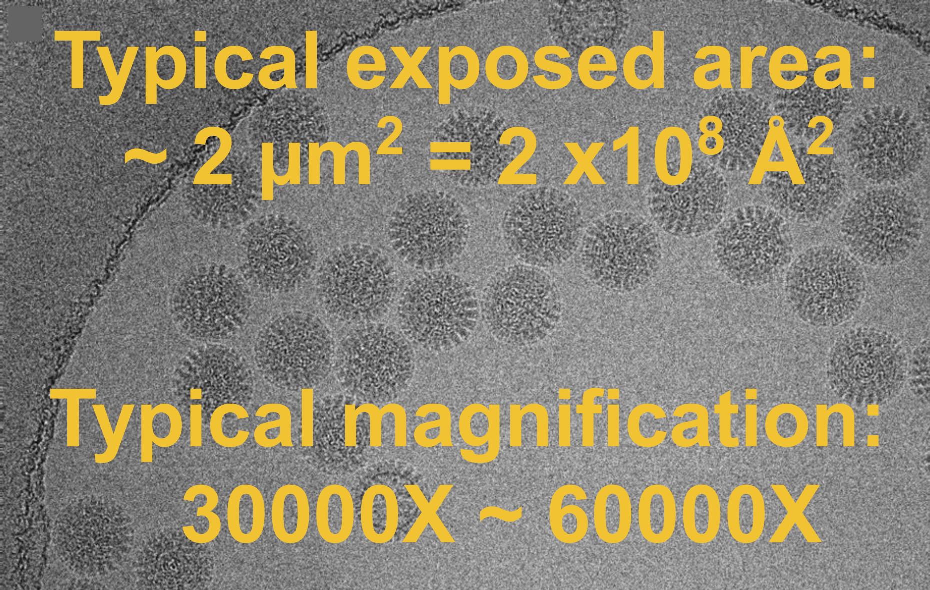
From: http://www.microbiologyinfo.com/differences-between-light-microscope-and-electron-microscope/



From: http://www.microbiologyinfo.com/differences-between-light-microscope-and-electron-microscope/

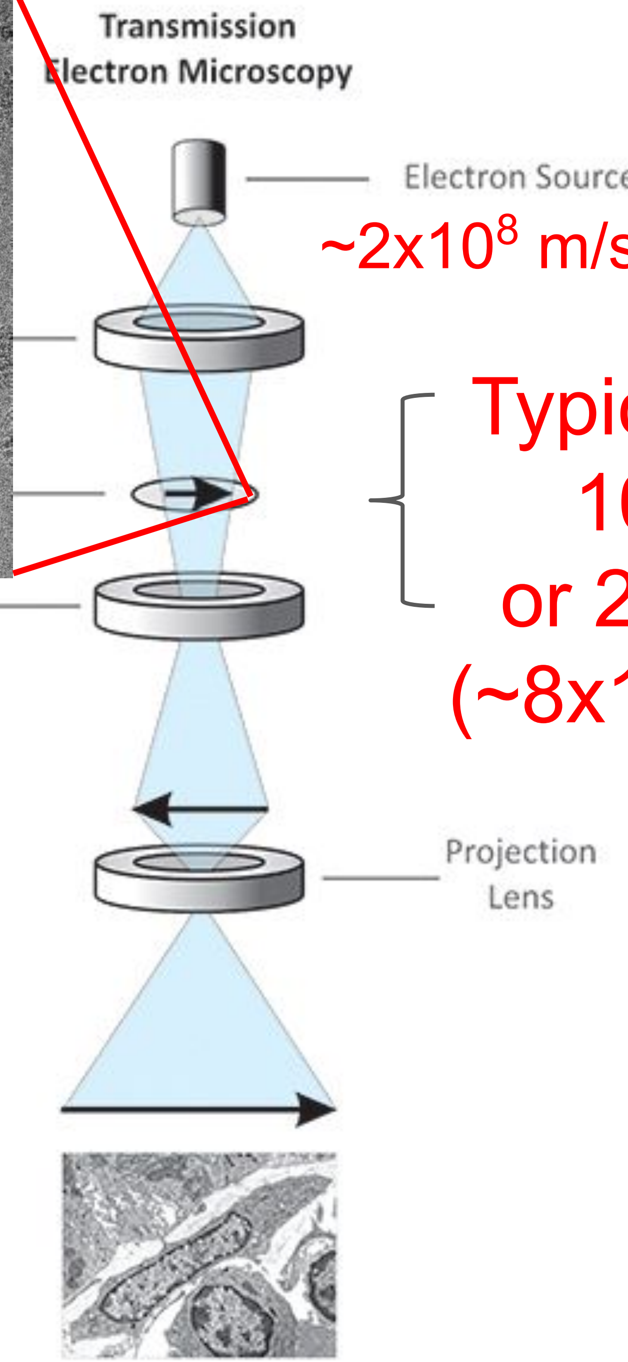


From: http://www.microbiologyinfo.com/differences-between-light-microscope-and-electron-microscope/



Typical exposed area:
~ 2 μm^2 = 2 x10^8 Å^2

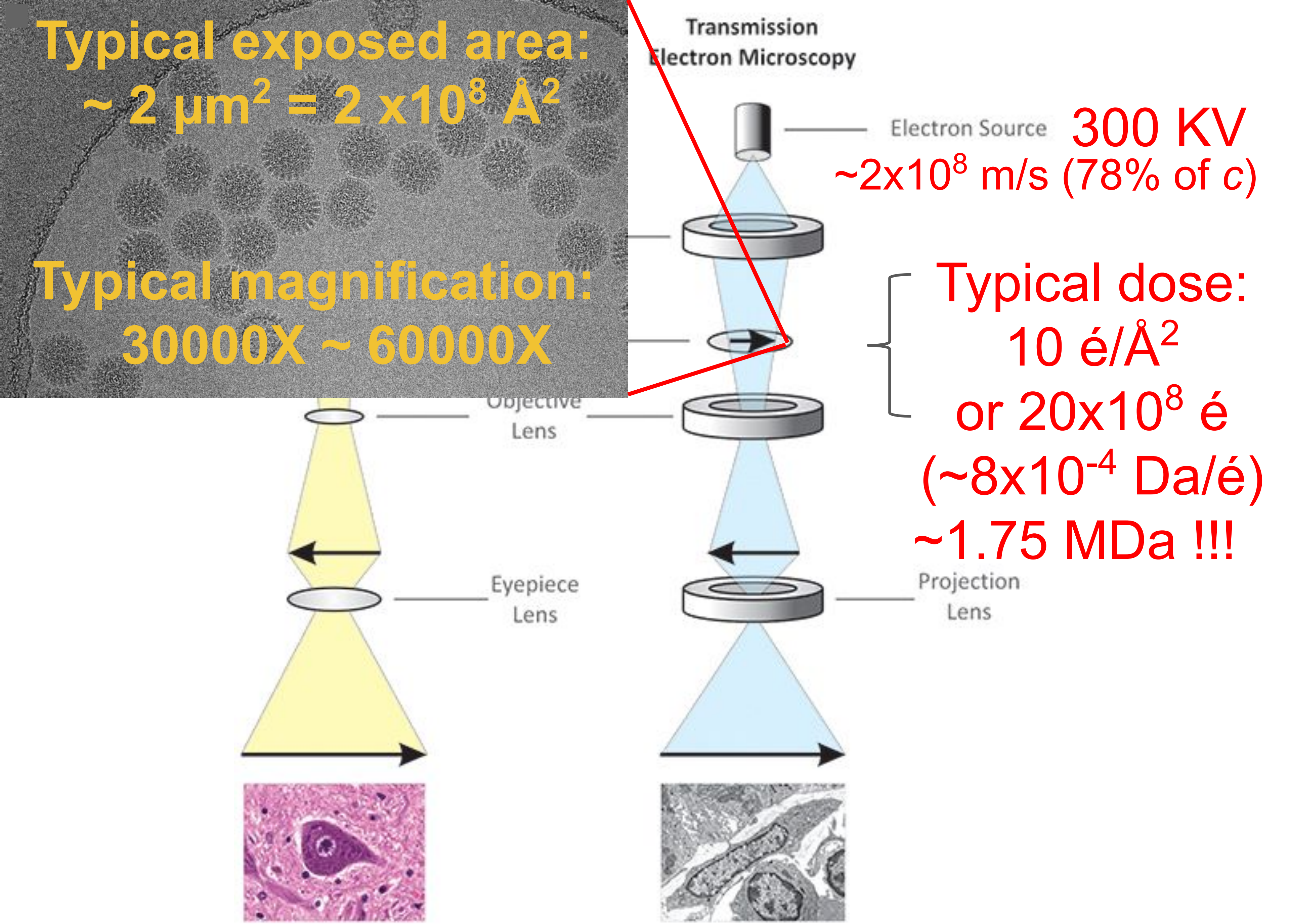
Typical magnification:
30000X ~ 60000X



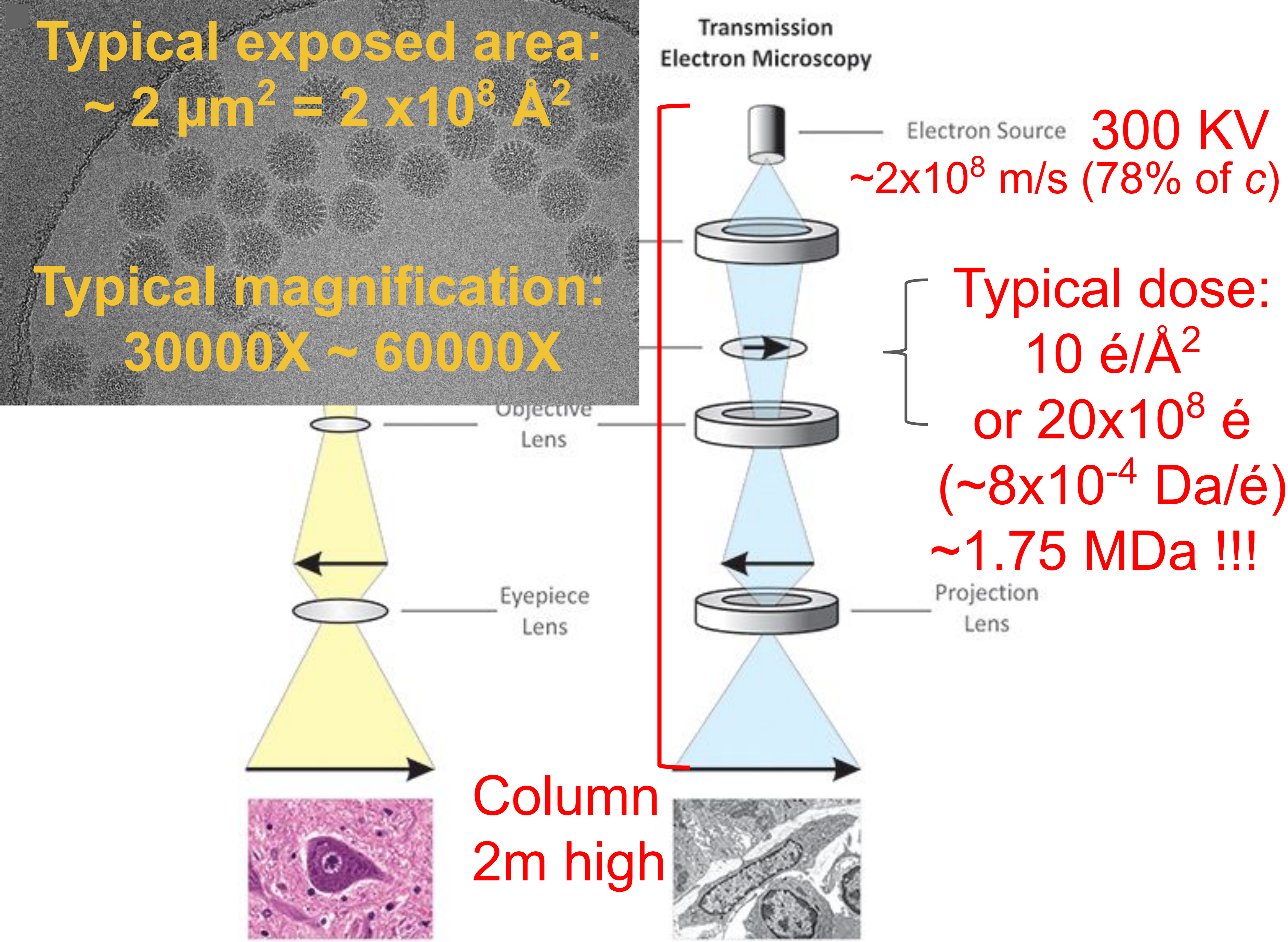
300 KV
~2x10^8 m/s (78% of *c*)

Typical dose:
10 é/Å^2
or 20x10^8 é
(~8x10^{-4} Da/é)

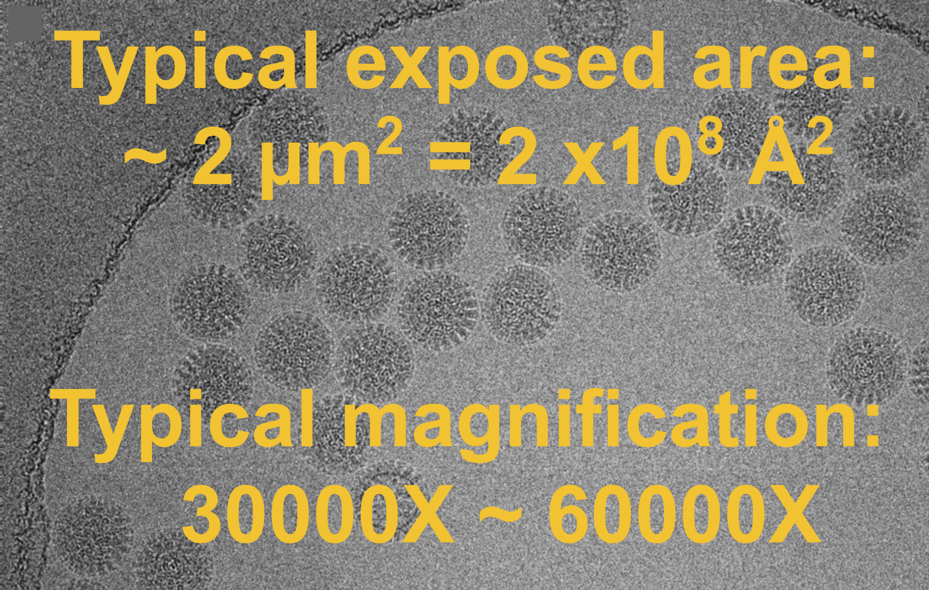
From: http://www.microbiologyinfo.com/differences-between-light-microscope-and-electron-microscope/

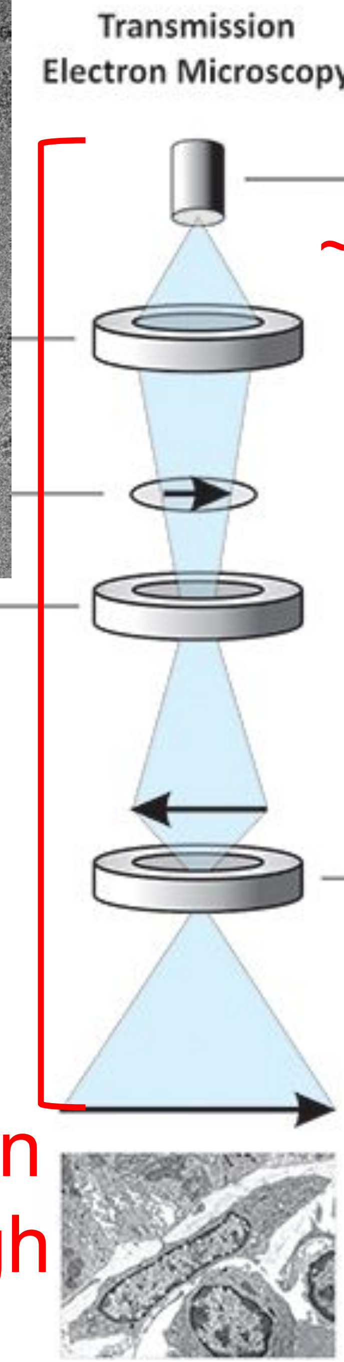


From: http://www.microbiologyinfo.com/differences-between-light-microscope-and-electron-microscope/



From: http://www.microbiologyinfo.com/differences-between-light-microscope-and-electron-microscope/





300 KV

~2×10^8 m/s (78% of *c*)

Typical dose:

10 é/Å^2

or 20×10^8 é

(~8×10^{-4} Da/é)

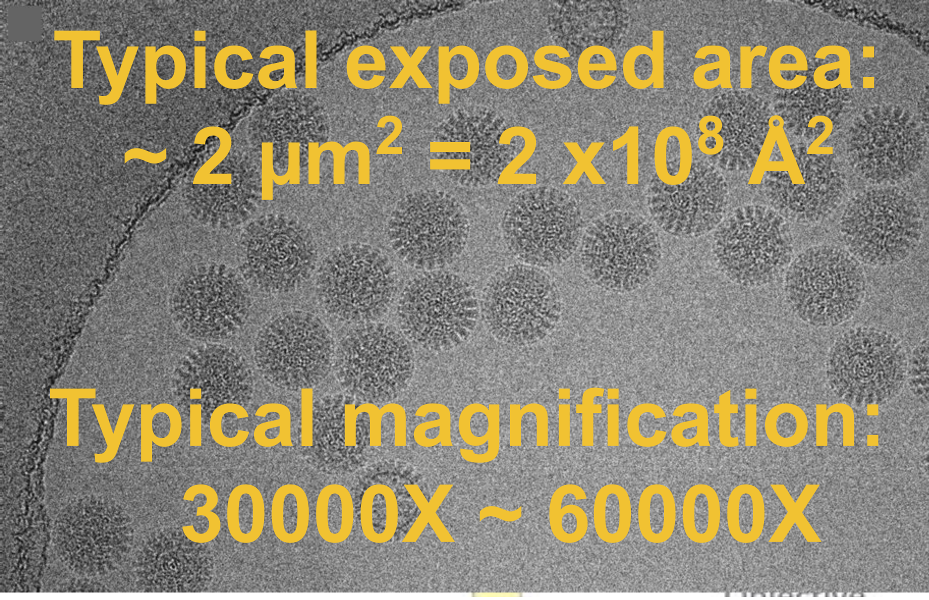
~1.75 MDa !!!

Column
2m high

2m / 2×10^8 m/s =
1×10^{-8} s

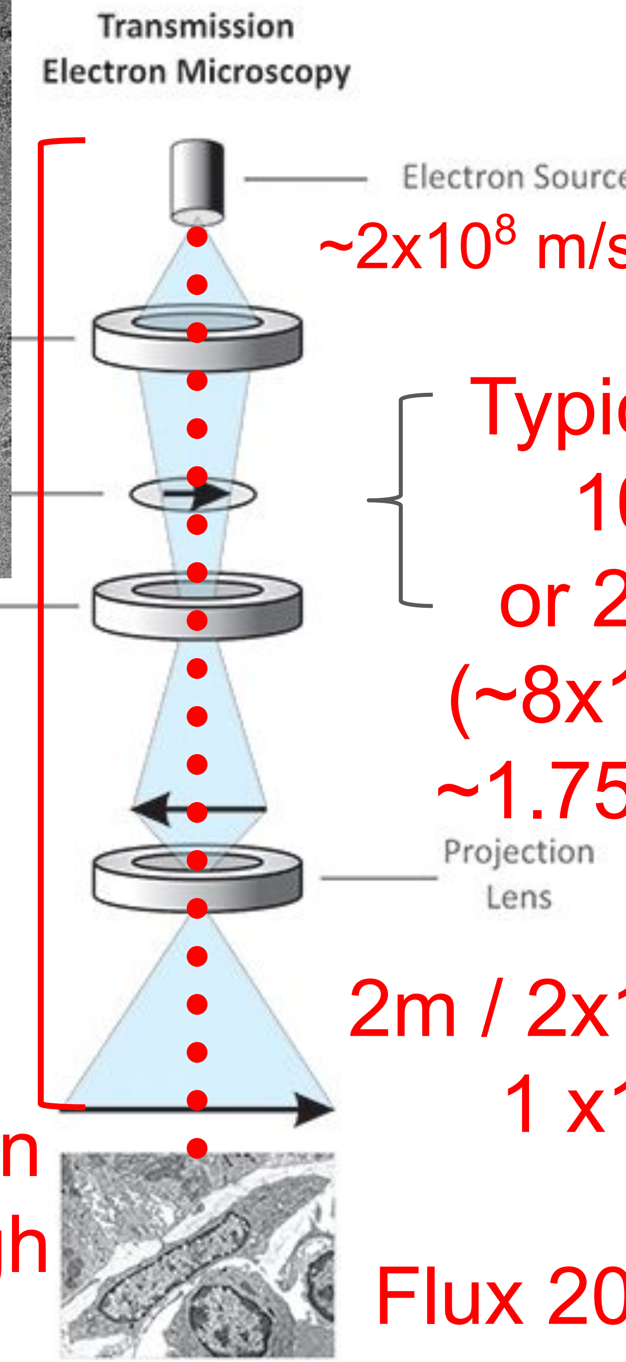
Flux 20×10^8 é/s

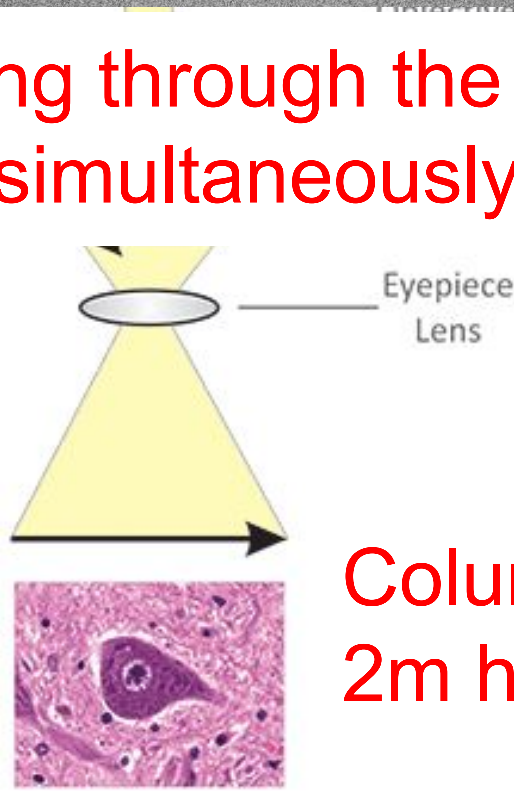
From: http://www.microbiologyinfo.com/differences-between-light-microscope-and-electron-microscope/



Typical exposed area:
~ 2 μm^2 = 2 x10^8 Å^2

Typical magnification:
30000X ~ 60000X





300 KV
~2x10^8 m/s (78% of *c*)

Typical dose:
10 é/Å^2
or 20x10^8 é
(~8x10^{-4} Da/é)
~1.75 MDa !!!

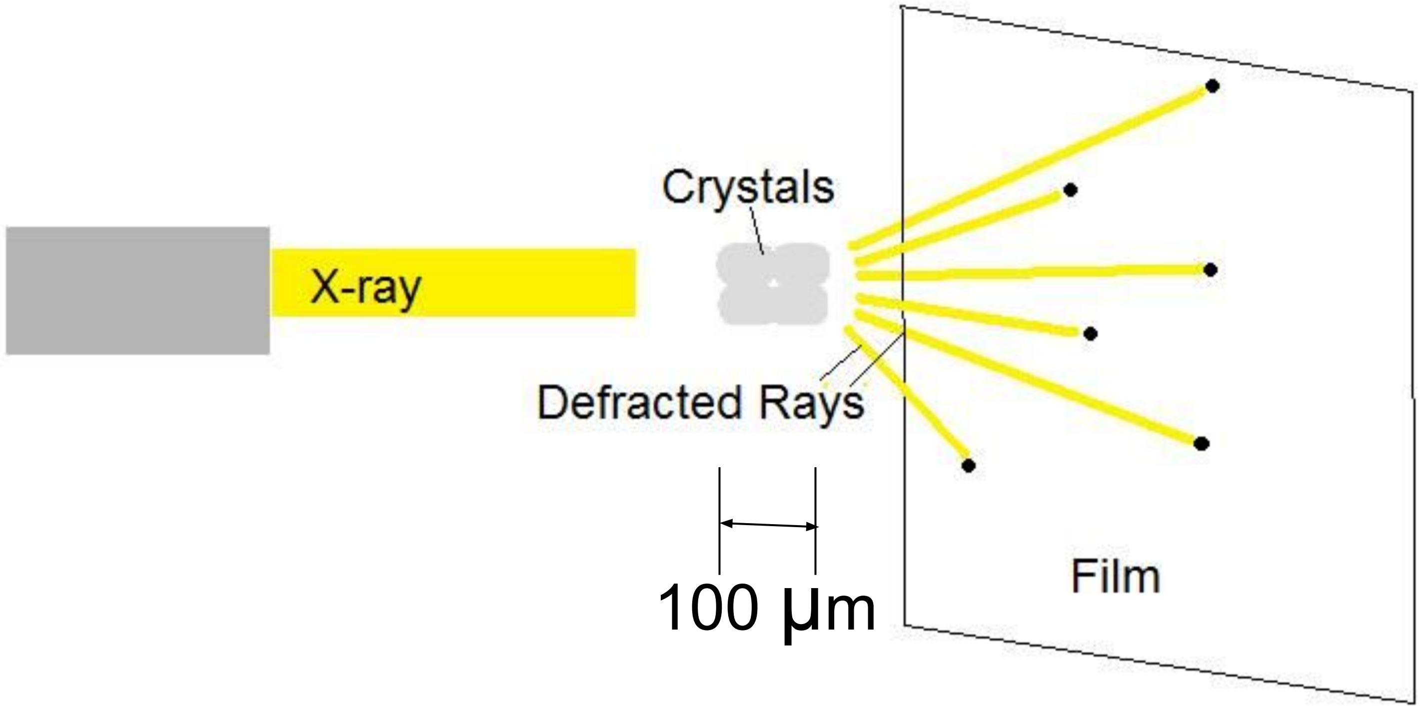
20 é flying through the column simultaneously

2m / 2x10^8 m/s =
1 x10^{-8} s

Column 2m high

Flux 20x10^8 é/s

From: http://www.microbiologyinfo.com/differences-between-light-microscope-and-electron-microscope/



at the detector level: 10^8 photons / s / mm^2

Detector surface: ~25220 mm^2

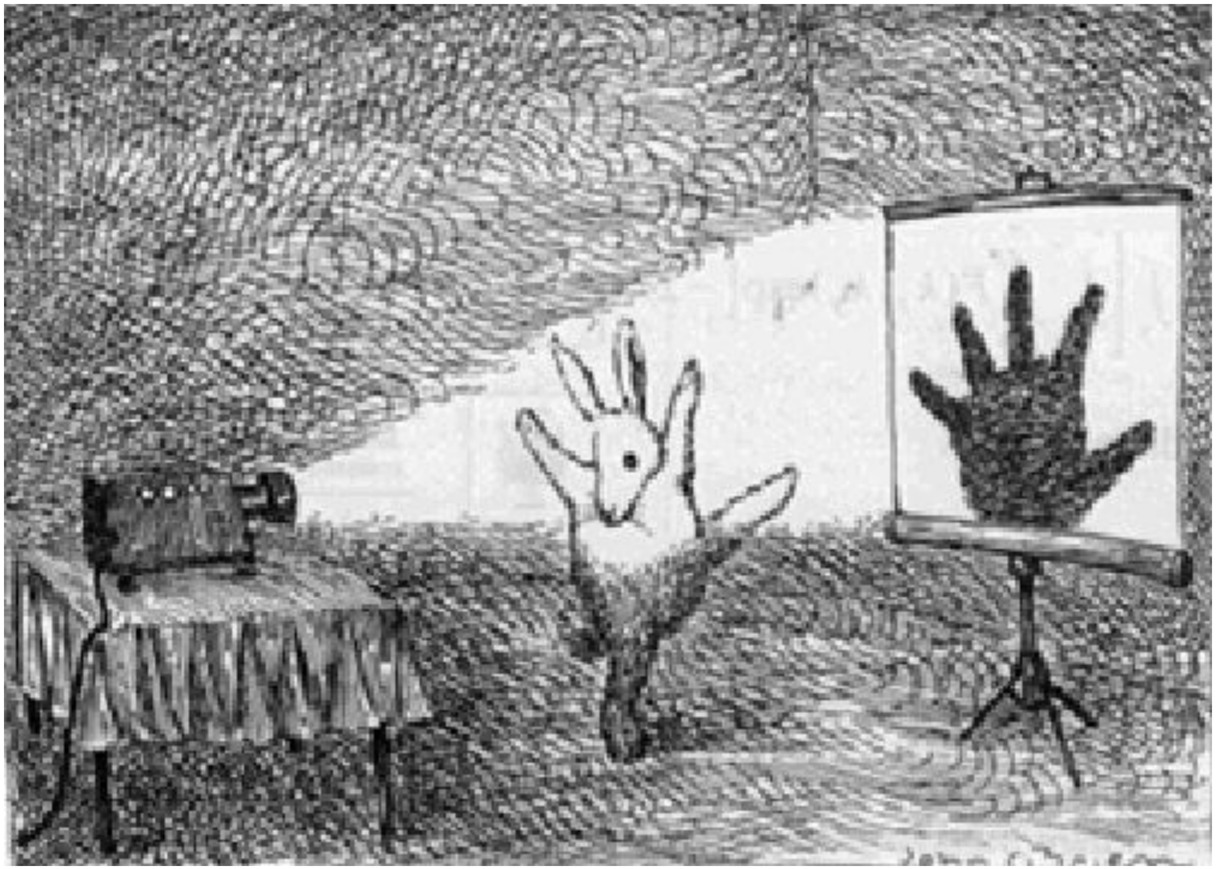
10^8 * 25220 = 2.522 * 10^{12} photons / s

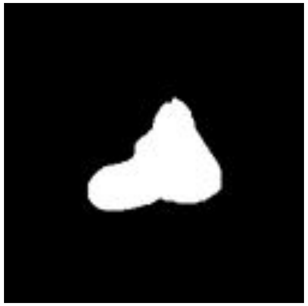
100 µm / 3*10^8 m/s = 3.3*10^{-14} s ~= 33 fentoseconds

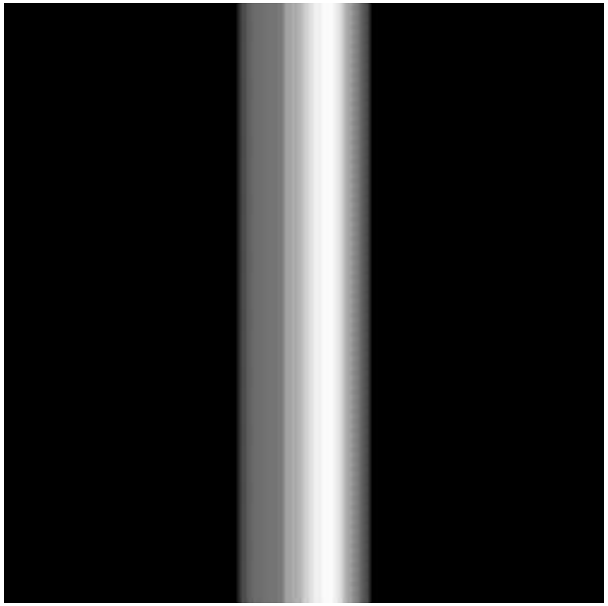
2.522*10^{12} photons/s * (3.3 *10^{-13} s) =

8.4 photons/100 µm (inside the crystal)

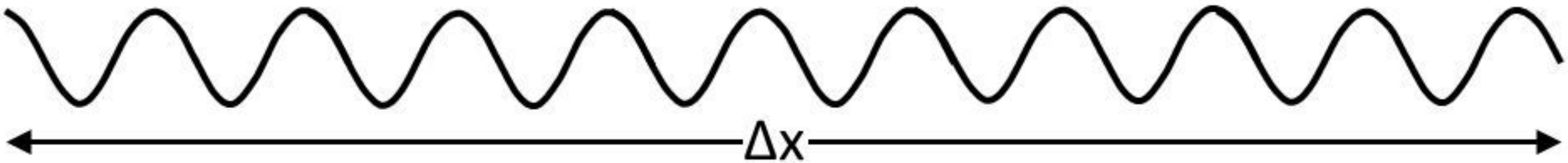
i.e. the photons are spaced by ~12 µm !



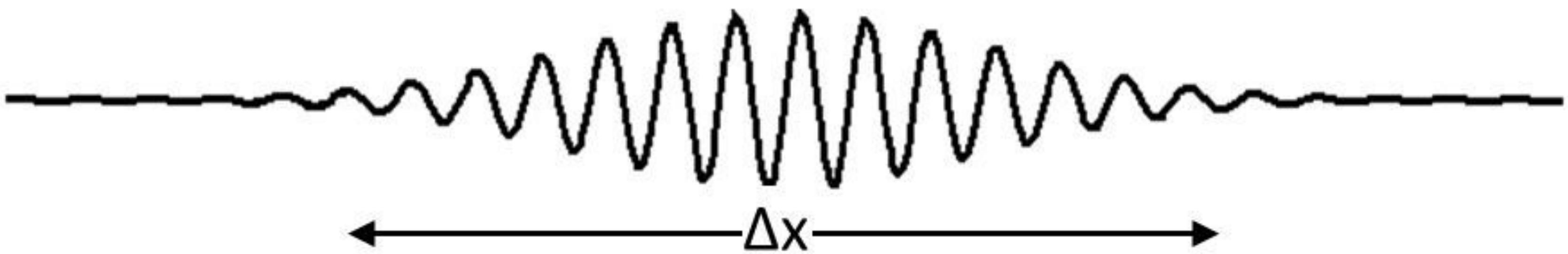




Uncertainty Principle



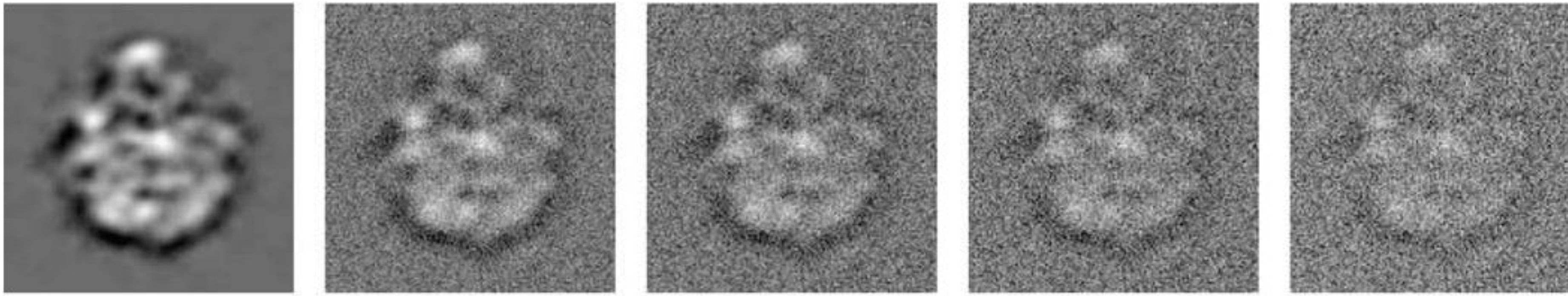
small Δp – only one wavelength



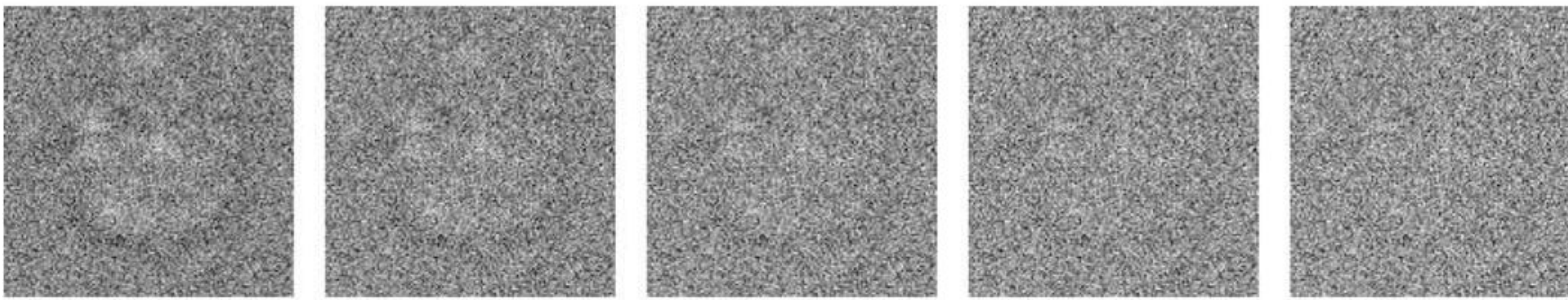
medium Δp – wave packet made of several waves

large Δp – wave packet made of lots of waves

Signal to noise ratio



(a) Clean (b) SNR=1 (c) SNR=1/2 (d) SNR=1/4 (e) SNR=1/8



(f) SNR=1/16 (g) SNR=1/32 (h) SNR=1/64 (i) SNR=1/128 (j) SNR=1/256

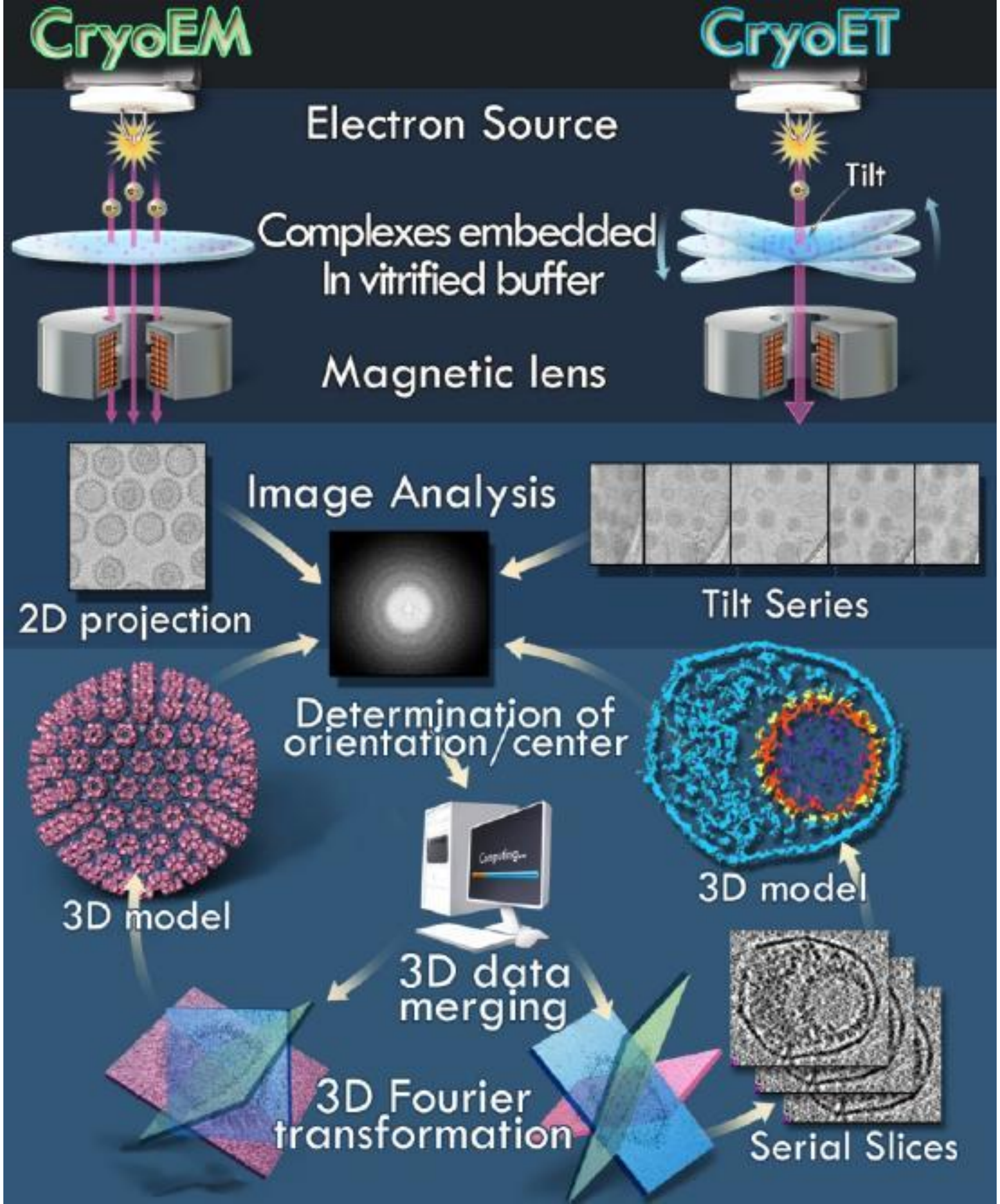
$$\mathrm{SNR} = \frac{\sigma^2_{\mathrm{signal}}}{\sigma^2_{\mathrm{noise}}}.$$

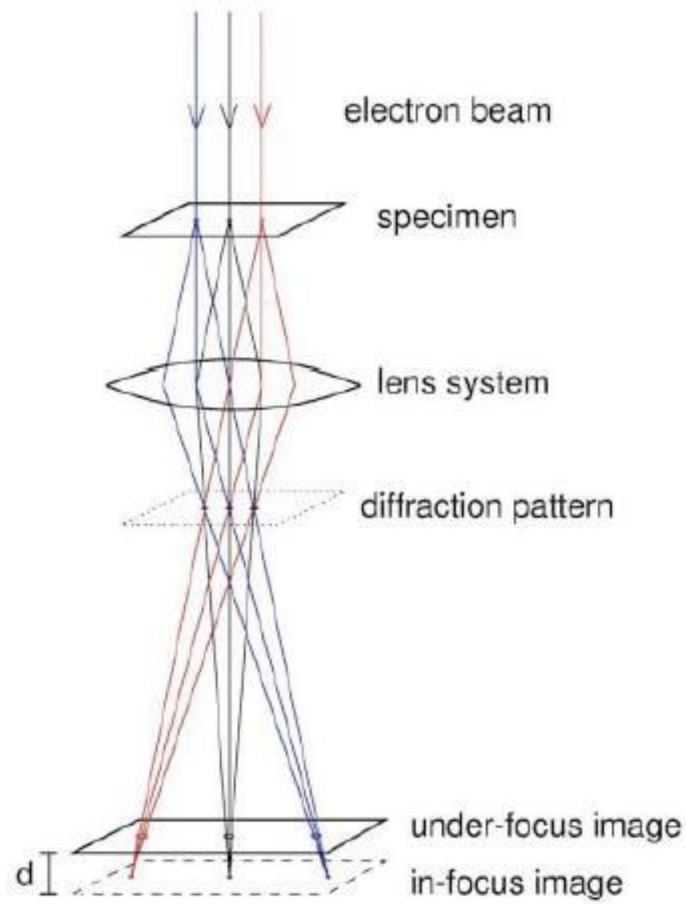
$$SSNR(r) = \frac{\sum_{r_i \in R} \left| \sum_{k_i} F_{r_i,k} \right|^2}{\frac{K}{K-1} \sum_{r_i \in R} \sum_{k_i} \left| F_{r_i,k} - \bar{F}_{r_i} \right|^2} - 1$$

From: https://www.semanticscholar.org/paper/Detecting-consistent-common-lines-in-cryo-EM-by-Singer-Coifman/639029f2c4af8d9f384d623695e3dcb43590fc02

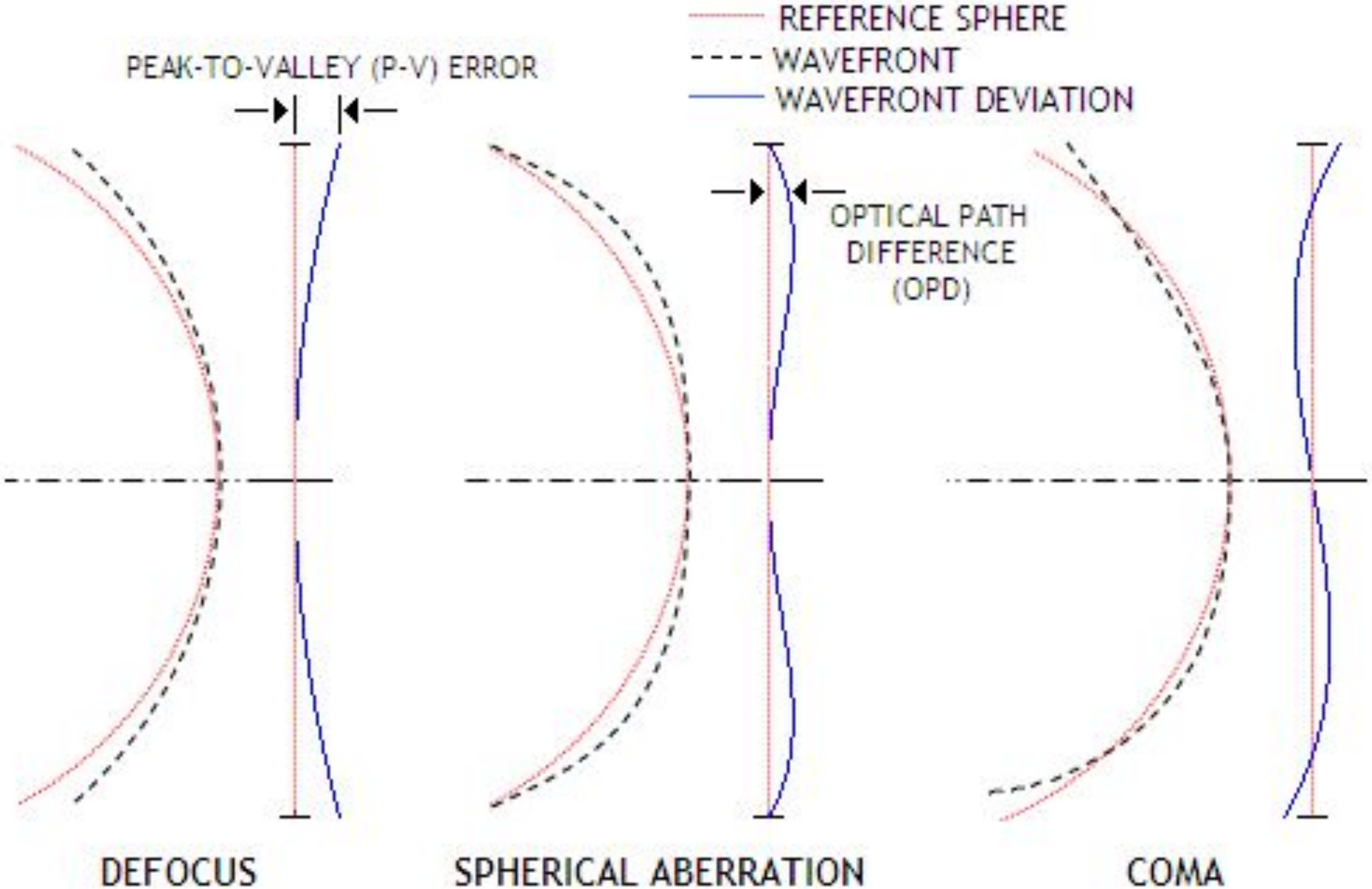
Definitions

1. *Information:* any entity or form that resolves uncertainty or provides some answer to some kind of question.
2. *Data:* something from which information can be extracted (or not).
3. *Image:* a visible impression containing information obtained by a camera, telescope, microscope, or other device.
4. *Digital image processing:* to perform operations on images using digital equipment (e.g. computers).
5. *Digital:* data represented as a finite sequence of finite discrete values.
6. *Operation:* any of various mathematical or logical processes (such as addition, multiplication, etc) of deriving one entity from others according to a rule.
7. *Digital image analysis:* process by which information is extracted from images.


CryoEM
CryoET
Electron Source
Tilt
Complexes embedded
In vitrified buffer
Magnetic lens
Image Analysis
2D projection
Tilt Series
Determination of
orientation/center
3D model
3D model
3D data
merging
3D Fourier
transformation
Serial Slices



Volkmann & Hanein, 2002



from: http://www.telescope-optics.net/aberrations.htm

Overiew

- Digital image representation
- Microscopes are linear time-invariant systems
- Signal processing and Fourier transform
- Point spread function
- Convolution and deconvolution
- Low, High and Band pass filtering
- Central slice theorem and 3D reconstruction
- Point-group symmetry (icosahedral)