IV Master in Biophysics
Universidad Autónoma de Madrid
Oct 26 – Nov 8/2006

Juan F. Poyatos

Stochastic dynamics

Evolutionary Systems Biology Lab

JUAN F. POYATOS
SPANISH NATIONAL CANCER CENTRE

http://bioinfo.cnio.es/~jpoyatos/

day II


0010100
001011001
001110011011
111100
01001
01011
100001
100 1
100G0AT
ATT01GCA0A
A0T01
ATTCG
TATGGATGG
TTA

Stochastic dynamics of gene expression, summary

- Genes are expressed by means of chemical reactions

- Chemical reactions are stochastic processes (collisions, etc)

- Gene expression is noisy: ***intrinsic*** noise (fluctuating reaction rates) TODAY
extrinsic noise (molecules involved in gene expression)

A simple model of gene expression, summary

$$\emptyset \xrightarrow{k} P,$$
$$P \xrightarrow{\delta} \emptyset,$$

Poisson process (birth and death)

- protein produced on average every 1/k seconds (birth)
- protein decays with rate δ (death)

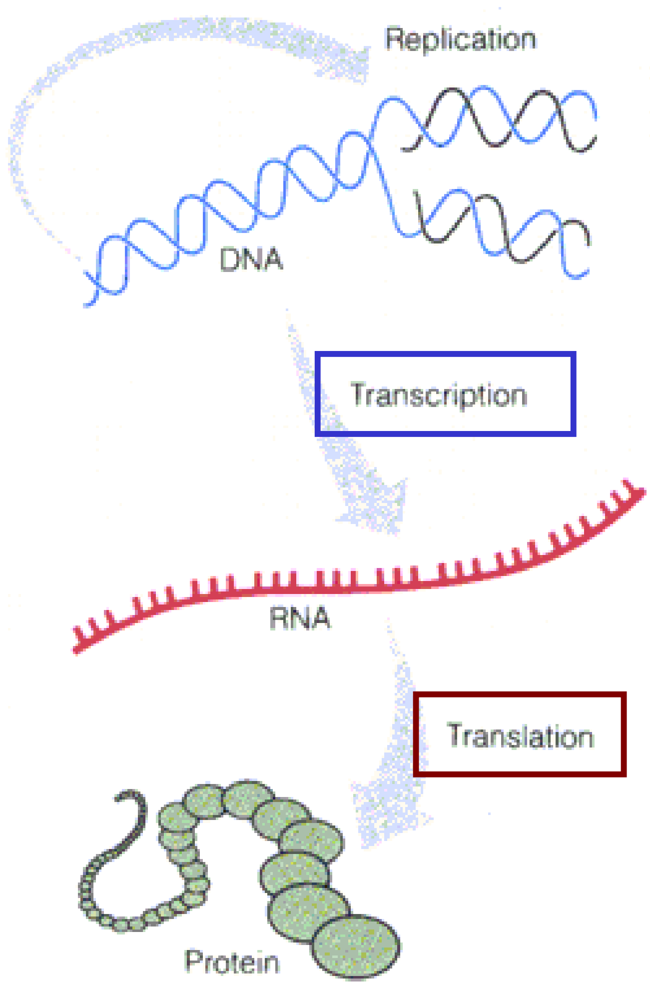
$$p_n = \frac{\langle n \rangle_{ss}^n}{n!} e^{\langle n \rangle_{ss}}$$

the steady state distribution is the Poisson Distribution

What is noise then? **CV**: $\eta_1 = \frac{\sigma}{\langle n \rangle}$ ($= 1/\sqrt{\langle n \rangle}$. Poisson distribution, noise increases as the number of molecules decreases)

Fano: $\eta_2 = \frac{\sigma^2}{\langle n \rangle}$ (= 1, Poisson distribution)

A more detail model of gene expression



Deterministic model

$$\frac{d[mRNA]}{dt} = k_R - \delta_R[mRNA]$$

$$\frac{d[P]}{dt} = k_P[mRNA] - \delta_P[P]$$

$$\emptyset \xrightarrow{k_R} mRNA,$$

$$mRNA \xrightarrow{\delta_R} \emptyset,$$

$$mRNA \xrightarrow{k_P} P,$$

$$P \xrightarrow{\delta_P} \emptyset,$$

P(k_R reaction) = $\mathbf{r}_k\, \delta t$

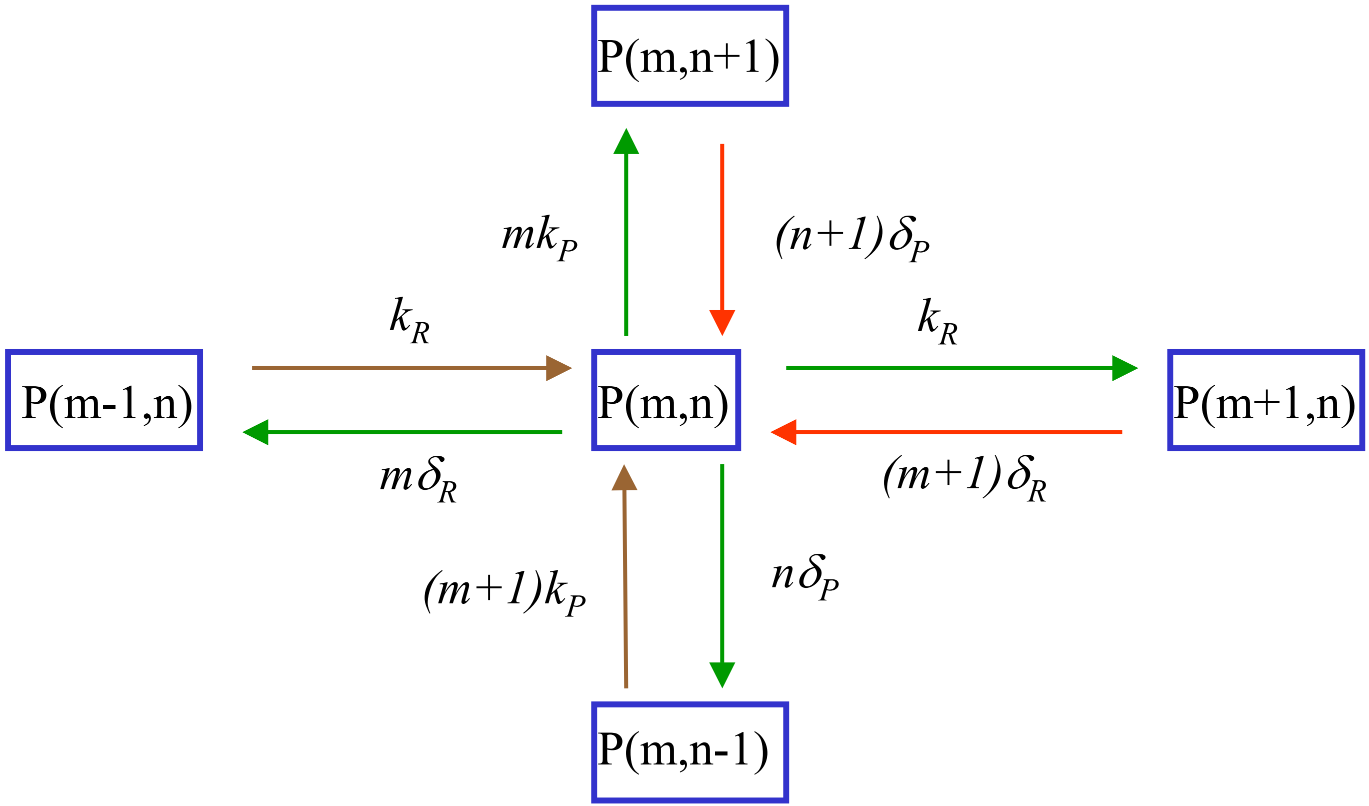
P(δ_R reaction) = $m_R\, \mathbf{r}_\delta\, \delta t$

P(k_P reaction) = $m_R\, \mathbf{r}_\kappa\, \delta t$

P(δ_P reaction) = $n_P\, \mathbf{r}_\delta\, \delta t$

How does the probability of having, say, m mRNA molecules and n *P* molecules, p(m,n), change with time?

r_κ's as k's to simplify notation
(this could also imply that $V = 1$)



$$
\begin{aligned}
\frac{dp_{m,n}}{dt} &= -p_{m,n}[m\delta_R + mk_P + k_R + n\delta_P] \\
&+ \ p_{m,n+1}(n+1)\delta_P + p_{m+1,n}(m+1)\delta_R \\
&+ \ p_{m,n-1}k_P m + p_{m-1,n}k_R
\end{aligned}
$$

Equation of the mean; emergence of deterministic laws

note first, a useful equation for a given function *f(n,m)*

$$\frac{d\langle f_{n,m}\rangle}{dt} = -\langle f_{n,m} m\rangle \delta_R - \langle f_{n,m} m\rangle k_P - \langle f_{n,m}\rangle k_R - \langle f_{n,m} n\rangle \delta_P + \langle f_{n-1,m} n\rangle \delta_P + \langle f_{n,m-1} m\rangle \delta_R + \langle f_{n+1,m} m\rangle k_P + \langle f_{n,m+1}\rangle k_R$$

thus, we get

$$\frac{d\langle m\rangle}{dt} = k_R - \delta_R \langle m\rangle$$

→ this is the equation the very same equation we obtained for the simple model, i.e., it implies steady state Poisson statistics for mRNA

$$\frac{d\langle n\rangle}{dt} = k_P \langle m\rangle - \delta_P \langle n\rangle$$

→ what kind of protein macroscopic steady state statistic characterizes protein dynamics?

we make use of the following equations ...

$$\frac{d\langle n^2\rangle}{dt} = -2\langle n^2\rangle\delta_P + \langle n\rangle\delta_P + 2\langle nm\rangle k_P + \langle m\rangle k_P$$

$$\frac{d\langle nm\rangle}{dt} = -\langle nm\rangle(\delta_P + \delta_R) + \langle m^2\rangle k_P + \langle n\rangle k_R$$

... to get the final expressions for the macroscopic statistics

$$\text{Fano Protein} = \frac{\langle n^2\rangle - \langle n\rangle^2}{\langle n\rangle} = 1 + \frac{k_P/\delta_R}{1+\delta_P/\delta_R} \approx \boxed{1 + \frac{k_P}{\delta_R}}$$

translation efficiency influences noise

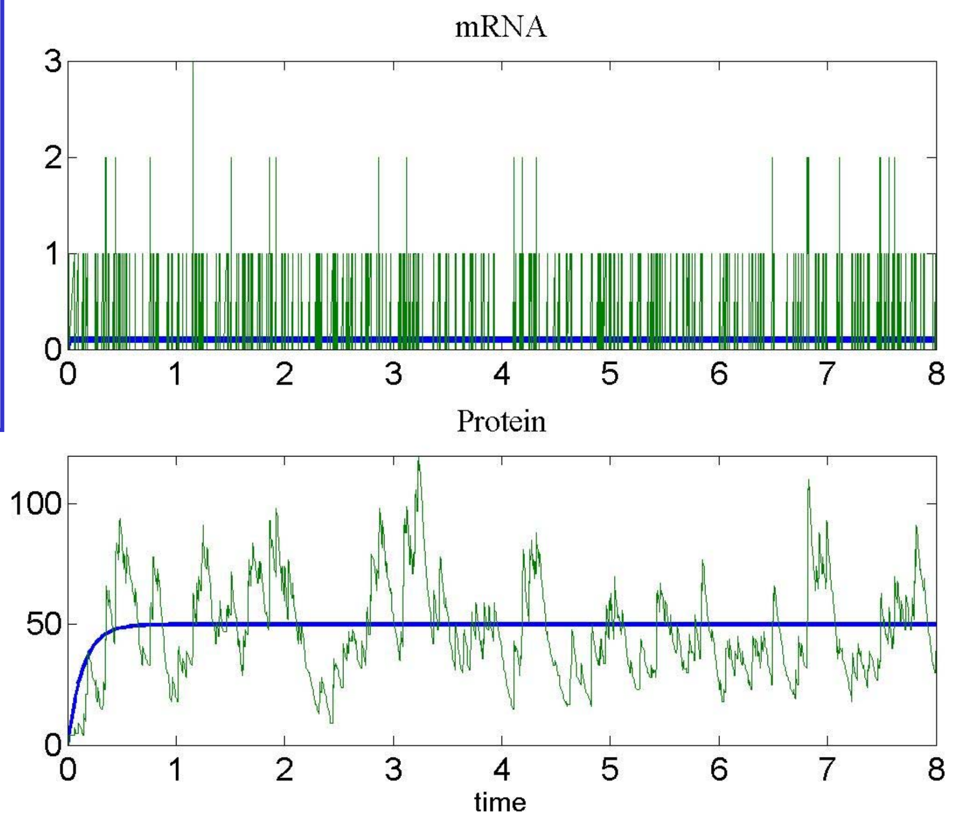
$$\text{Fano mRNA} = \boxed{1}$$

transcription efficiency does not influence noise

protein half-lifetime ~ hours
mRNA half-lifetime ~ minutes
thus

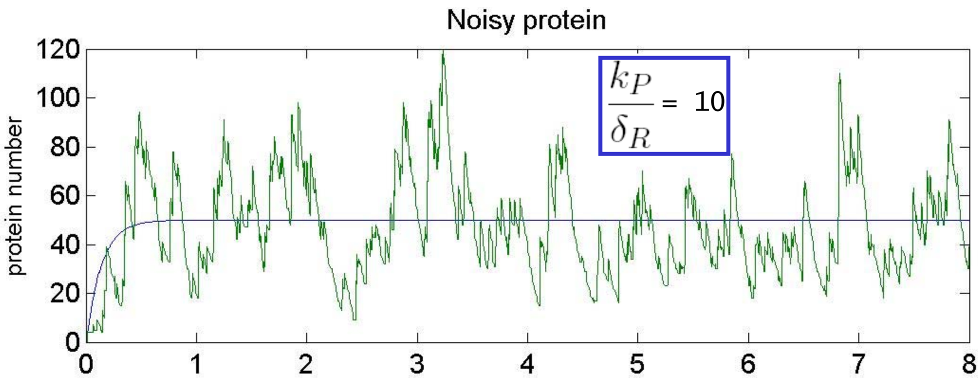
$t_{1/2} = \log 2/\delta$ and $\delta_P \ll \delta_R$

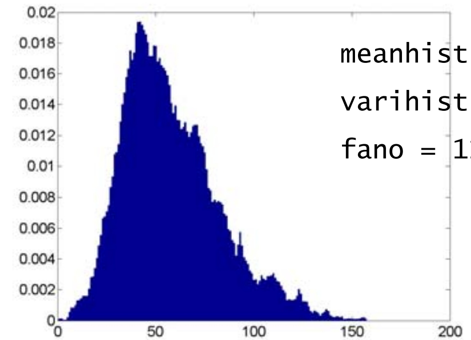
```
% .. code2stoch.m
% .. more detail gene expression stochastic and
deterministic

clear all
kR = .01;       % .. []/s
deltaR = .1;    % .. 1/s
kP = 10*deltaR; % .. 1/s
deltaP = .002   % .. 1/s

% .. stochastic eqs. Gillespie's algorithm
P = [0 0];
Pstochastic = P;
tmax = 8*60*60;      % .. hours
t = 0;
tspan = t;
```



```
while t < tmax
    % .. a's
    a = [kR, deltaR*P(1),kP*P(1),deltaP*P(2)];
    a0 = sum(a);

    % .. determine time of next reaction
    r1 = rand;
    tau = -log(r1)/a0;
    t = t + tau;
    % .. determine nature of next reaction
    r2 = rand;
    acumsum = cumsum(a)/a0;
    chosen_reaction = min(find(r2 <= acumsum));

    if chosen_reaction == 1;
        P(1) = P(1) + 1;
    elseif chosen_reaction == 2;
        P(1) = P(1) - 1;
    elseif chosen_reaction == 3;
        P(2) = P(2) + 1;
    else
        P(2) = P(2) - 1;
    end

    tspan = [tspan,t];
    Pstochastic = [Pstochastic;P];
end
```

```
% .. deterministic eqs.
P0 = [0,0];
options = [];
[t P] = ode23(@code2equations,tspan,P0,options,kR,deltaR,kP,deltaP);

% .. plot
subplot(211);plot(t/60/60,P(:,1),t/60/60,Pstochastic(:,1))
axis([0 tmax/60/60 0 max(Pstochastic(:,1))]);title('mRNA');
subplot(212);plot(t/60/60,P(:,2),t/60/60,Pstochastic(:,2))
axis([0 tmax/60/60 0 max(Pstochastic(:,2))]);title('Protein')
```

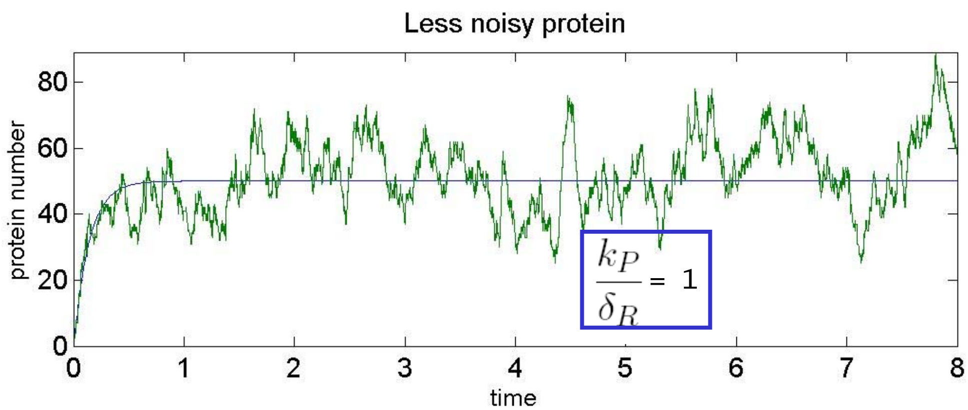


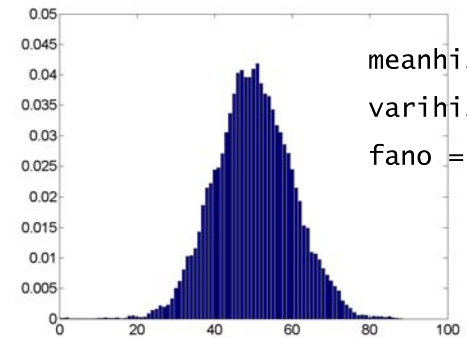


meanhist = 57.1261

varihist = 634.6932

fano = 11.1104





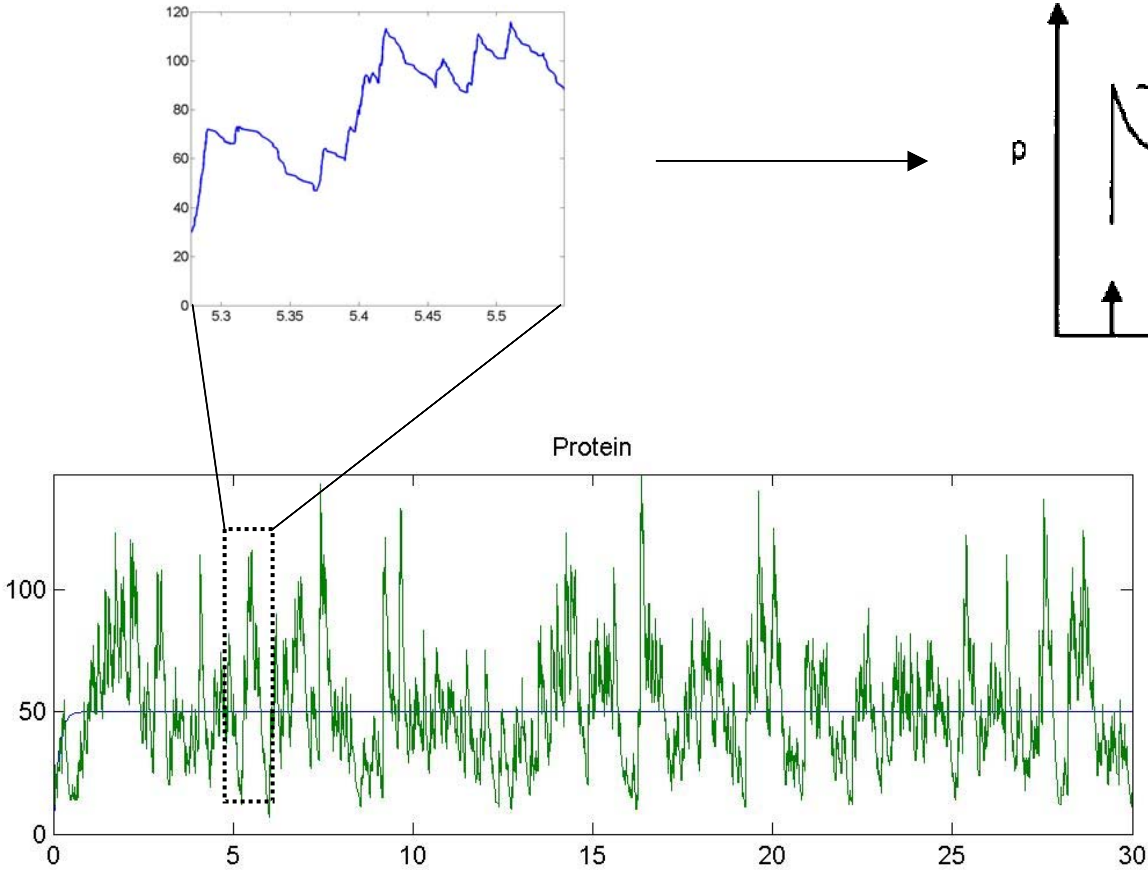
meanhist = 50.0542

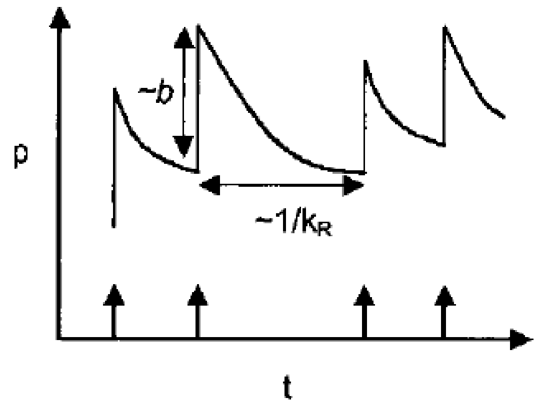
varihist = 100.2078

fano = 2.0020

"Random bursts model"

$$b = \frac{k_P}{\delta_R}$$





Master equations and gene expression

- Genes are generally regulated by complex nonlinear functions. Analitical studies become difficult.

- Two types of approximation methods

1) Numerical Simulation → Gillespie's algorithm

2) Perturbation Methods → Langevin equations, ...

Langevin equations

- Developed originally to the study of Brownian motion

- Alternative mathematical framework to that of the Master Equation

- Better suited for an intermediary ("not very noisy") regime

- Based on adding explicitely noise terms to the deterministic (macroscopic) equations

- Need to characterize the noise distribution of the added noise

$$\frac{d[mRNA]}{dt} = k_R - \delta_R[mRNA] + \underline{\xi_R}$$
$$\frac{d[P]}{dt} = k_P[mRNA] - \delta_P[P] + \underline{\xi_P}$$

ξ_R, ξ_P added stochastic variables.

These equations are fully specified when the probability distributions for the stochastic variables are given.

Valid to describe an intermediate situation where fluctuations are important even though the number of particles is big enough.

what are the properties of ξ ?

- we would like to know mean, variance, characteristic fluctuation times, ...

Fluctuation time:
we can ask how much correlates the variation of ξ with respect to its mean:
autocorrelation function

$$
\begin{aligned}
C_\xi(t_1, t_2) &= \Big\langle \big[\xi(t_1) - \langle \xi(t_1) \rangle\big] \big[\xi(t_2) - \langle \xi(t_2) \rangle\big] \Big\rangle \\
&= \langle \xi(t_1)\xi(t_2) \rangle - \langle \xi(t_1) \rangle \langle \xi(t_2) \rangle
\end{aligned}
$$

at equal times (t_1=t_2) we recover the variance.

Often there exists a characteristic time τ_c for which $C_\xi(t_1, t_2) = 0$.

τ_c is known as the autocorrelation time

White noise

Langevin originally applied to brownian motion: no reason why thermal fluctuations should favour a particular reaction:

$\xi_R \quad \xi_P$ defined such that $\langle \xi(t) \rangle = 0$

thus, $C_\xi(t_1, t_2) = \langle \xi(t_1))\xi(t_2)\rangle$

collision time is faster than time-scale of change of molecule numbers, noise is uncorrelated

$$C_\xi(t_1, t_2) = \Gamma e^{\frac{-(t_1 - t_2)}{\tau}}$$

noise strength, variance at equal times

$\tau \to 0$

very small autocorrelation times: $C_\xi(t_1, t_2) = \Gamma \delta(t_1 - t_2)$

White noise:

- noise variable with *zero* autocorrelation time

- white?

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}\langle\xi(t))\xi(t+\tau)\rangle d\tau = 1$$

all frequencies contribute equally

Color noise

- noise variable with *finite* autocorrelation time

Stochastic differential equations are very irregular, modified numerical methods to solve them: one to the simplest Euler-Maruyana method