# Probabilistic Model Checking

Marta Kwiatkowska
Gethin Norman
Dave Parker


University of Oxford

## Part 6 – CTMC Case Studies

# Overview

- Introduce two real-world examples
  - models are continuous-time Markov chains (CTMCs)
  - demonstrate a broad range of quantitative analyses possible with PRISM
- Dynamic power management
  - application domain of growing importance
- Biological systems
  - collaboration with experimental biologists
  - model described in stochastic pi-calculus as well as probabilistic reactive modules
  - predict outcome of experiments
- See PRISM web page for more…

# Power management

- Power management
  - controls power consumption in battery-operated devices
  - savings in power usage translate to extended battery life
  - important for portable, mobile and handheld electronic devices

- System level power management
  - manages various system devices for power optimisation
  - system components manufactured with several power modes
  - e.g. disk drive has: active, idle, standby, sleep, …
  - modes can be changed by the operating system through APIs
  - exploits application characteristics
  - needs to be implemented at the O/S level

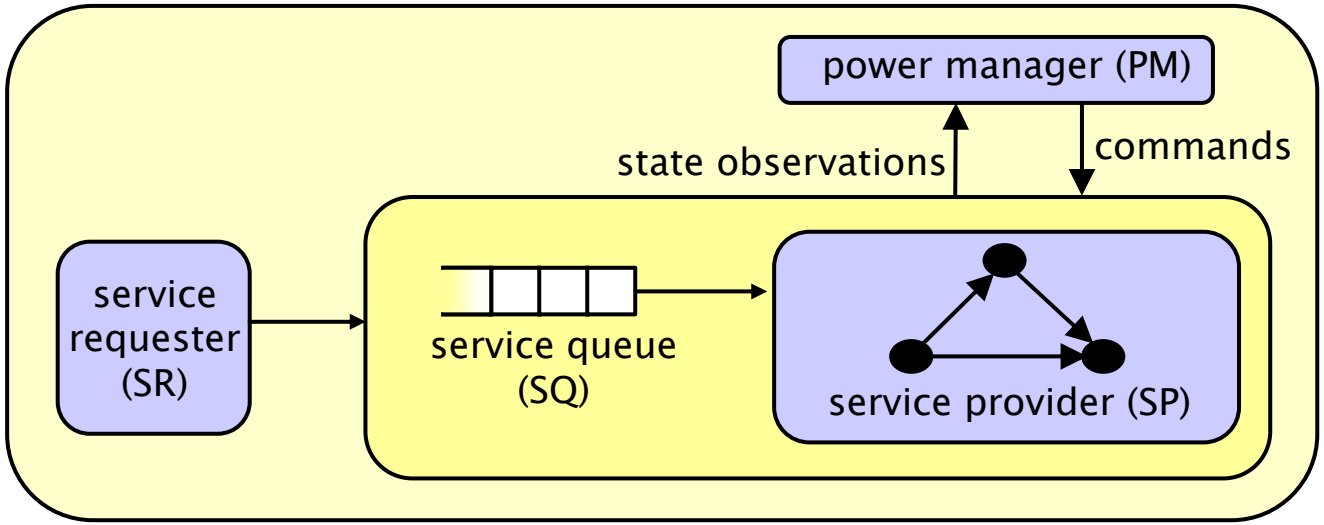
# Dynamic Power Management (DPM)

- DPM make optimal decisions at runtime based on:
  - dynamically changing system state
  - workload
  - performance constraints

- Stochastic optimal control strategies for DPM
  - construct a mathematical model of the system in PRISM
  - transition times modelled with exponential distributions
  - formulate stochastic optimisation problems
    e.g. “optimise av. energy usage while av. delay below k”
  - create stochastic strategies by solving optimisation problem
    (exported to Maple for solution externally)
  - analyse strategies in PRISM

# DPM – The system model

- Service requester (generates the service requests)
- Service provider (provides service to the requests)
- Service queue (buffers the requests)
- Power manager (monitors the states of the SP and SQ and issues state-transition commands to the SP)



# Fujitsu disk drive – The PRISM model

- 4 state Fujitsu disk drive: busy, idle, standby and sleep
- Policies:
  - minimize the average power consumption
  - constraint on the average queue size
- Reward structure "power" (power consumption)
  - state rewards: the av. power consumption of SP in the state
  - transition rewards: energy consumed when SP changes state
- Reward structure "queue" (queue size)
  - state rewards: current size of the queue
- Reward structure "lost" (lost requests)
  - transition rewards: assign 1 to transitions representing the arrival of a request in a state where the queue is full

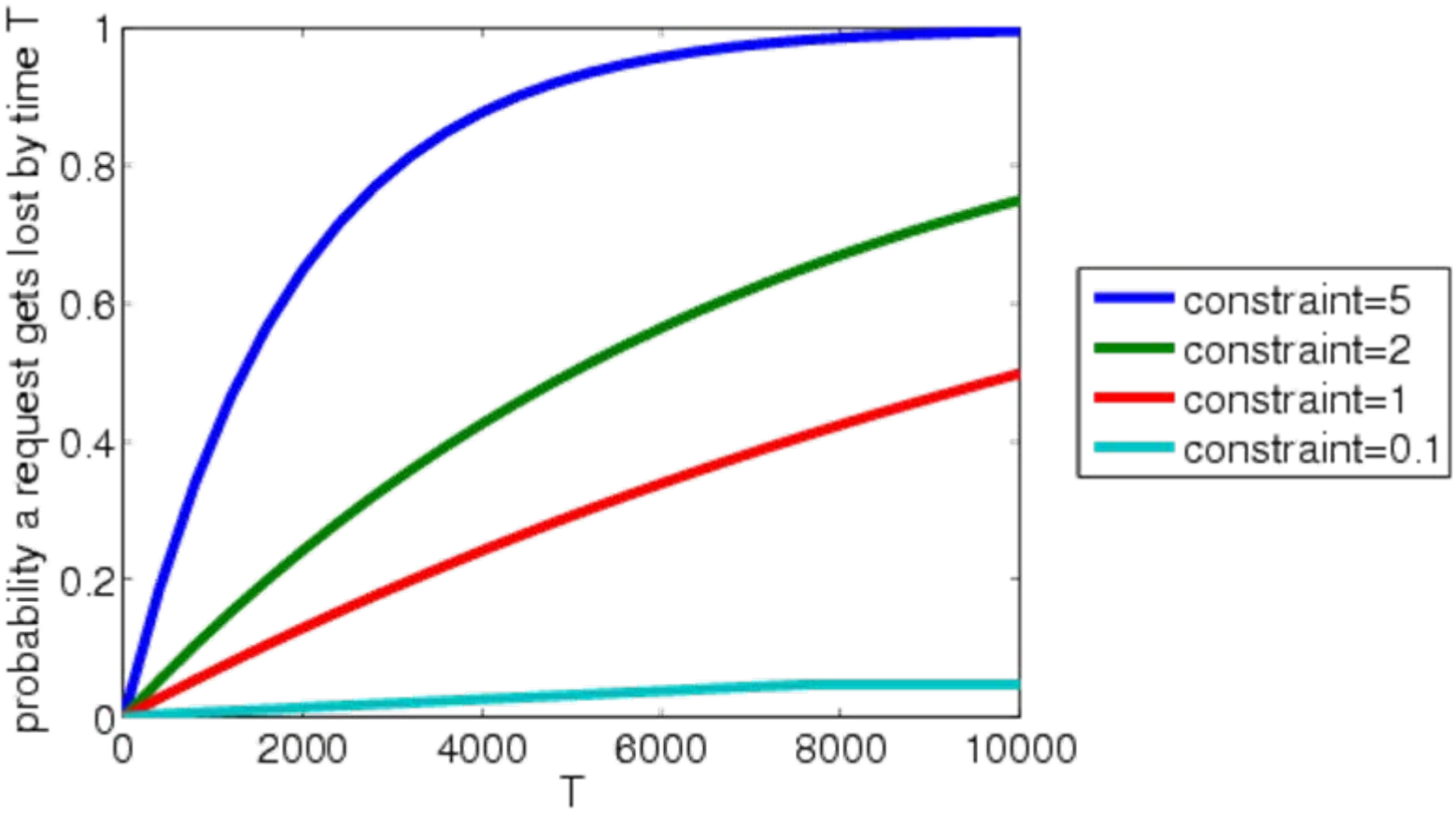
# Fujitsu disk drive – Properties

- Selection of properties checked with PRISM

- Probability that queue size becomes $\geq$ M by time t
  - $P_{=?}[F^{\leq t}\ (q \geq M)]$
- Probability that at least M requests get lost by time t
  - $P_{=?}[F^{\leq t}\ (\text{lost} \geq M)]$
- Expected queue size at time t
  - $R_{\{\text{“queue”}\}=?}[I^{=t}]$
- Expected power consumption by time t
  - $R_{\{\text{“power”}\}=?}[C^{\leq t}]$
- Long run average number of requests lost
  - $R_{\{\text{“lost”}\}=?}[S]$

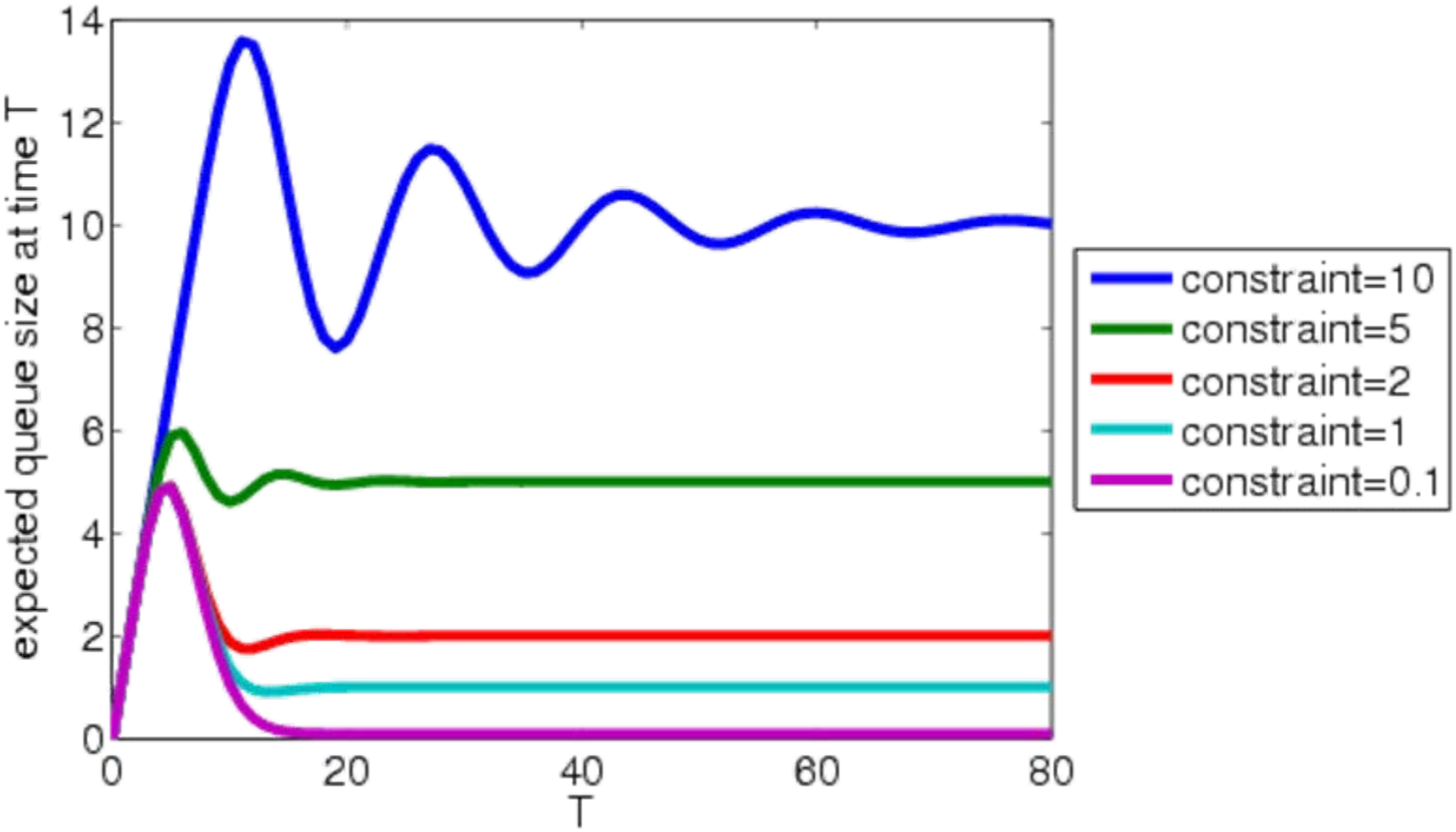
# Fujitsu disk drive – PRISM results

- Probability M requests lost by time t $P_{=?}[F^{\leq t} (lost \geq M)]$



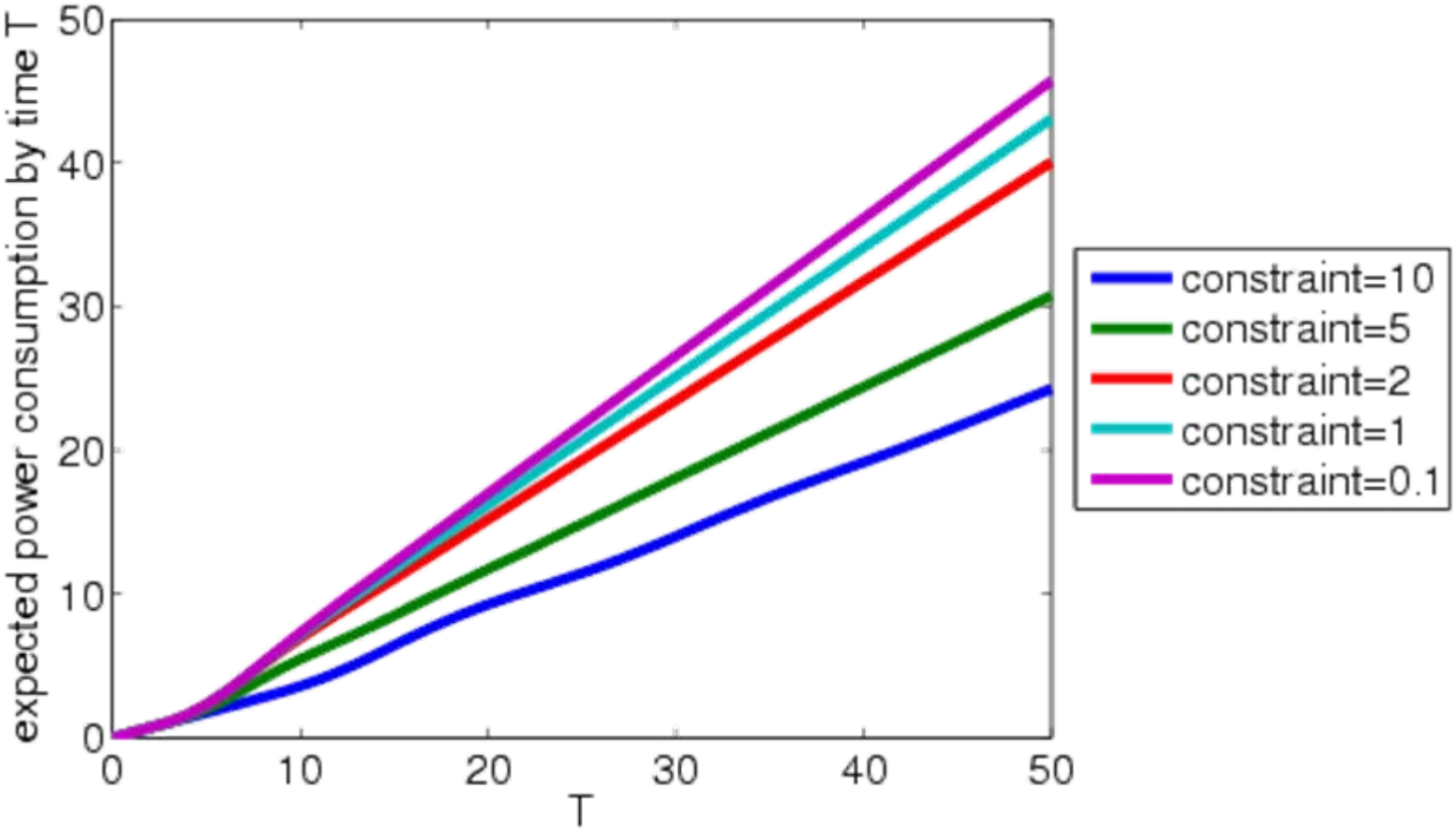
# Fujitsu disk drive – PRISM results

- Expected queue size at time t $R_{\{\text{“queue”}\}=?}[I^{=t}]$



# Fujitsu disk drive – PRISM results

- Expected power consumption by time t $R_{\{\text{“power”}\}=?}[C^{\leq t}]$



# Overview

- Introduce two real-world examples
  - models are continuous-time Markov chains (CTMCs)
  - demonstrate a broad range of quantitative analyses possible with PRISM
- Dynamic power management
  - application domain of growing importance
- Biological systems
  - collaboration with experimental biologists
  - model described in stochastic pi-calculus as well as probabilistic reactive modules
  - predict outcome of experiments
- See PRISM web page for more…

# Biological systems

- Networks of subsystems
  - organisms, cells, molecules, …
- Interaction
  - governed by rules
  - causes transformations
- Evolution
  - continuous and discrete dynamics
- Mobility
  - motion in space and time, re-configurability, …
- Stochastic behaviour
  - unpredictability, noise, …

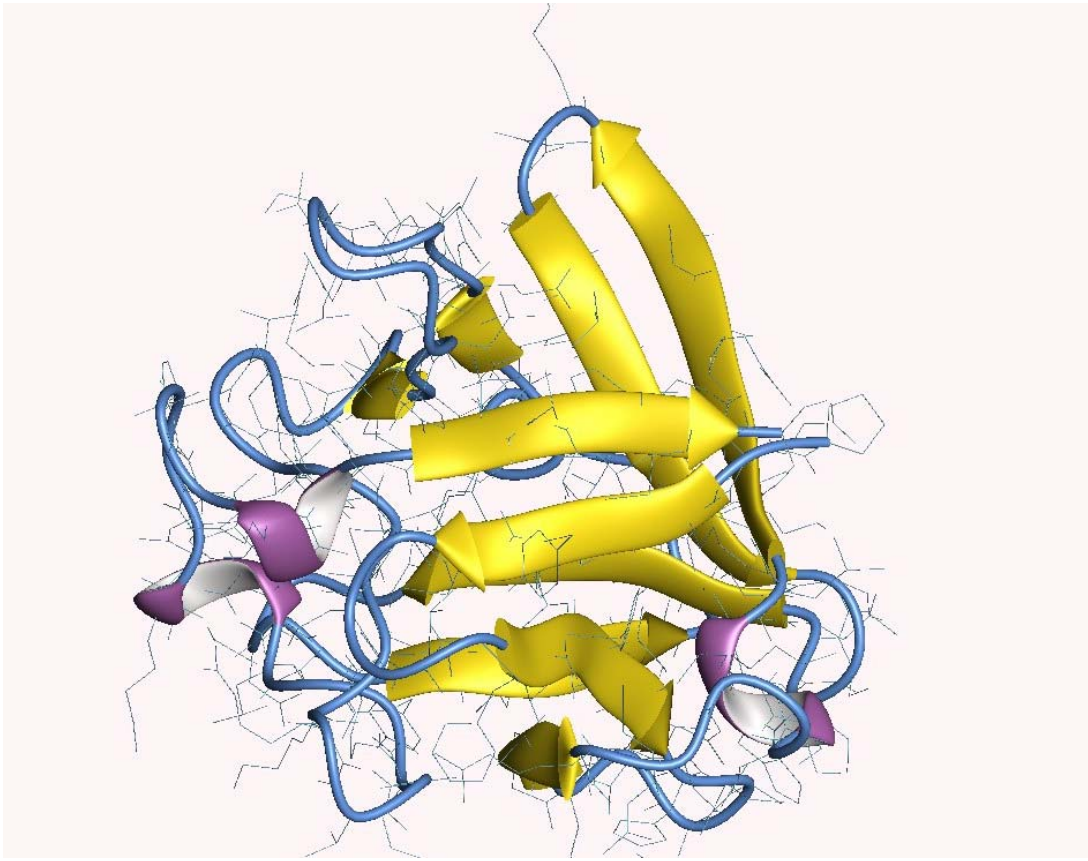
Not unlike computers, networks and the Internet…

Reuse methods for systems biology?

- Propose to use process calculi to model biological processes [Regev, Shapiro, Cardelli, …]

# Modelling signalling pathways

- Focus on
  - networks of molecules
  - interaction
  - continuous & discrete dynamics
- Rather than
  - geometry
  - structure
  - sequence



Google images: Human FGF, http://160.114.99.91/astrojan/prot1t.htm

# Modelling frameworks

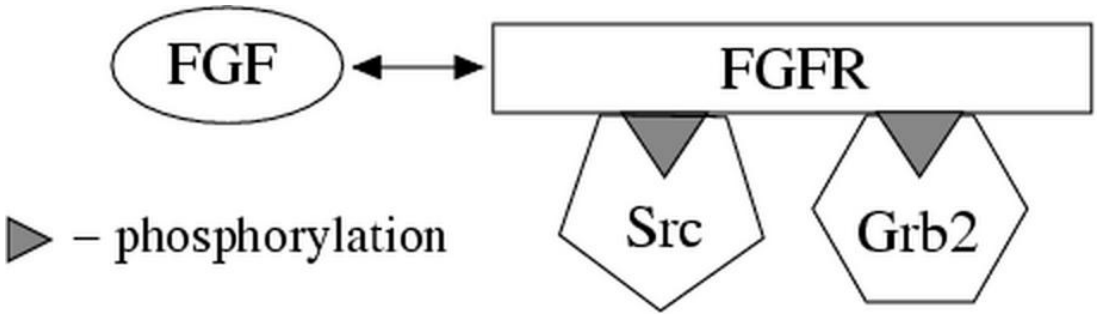
- Assume wish to model mixture of molecules
  - N different molecular species, interact through reactions
  - fixed volume V (spatially uniform), constant pressure and temperature
- Continuous deterministic approach
  - approximate the number of molecules in V at time t by a continuous function, if large numbers of molecules
  - obtain ODEs (ordinary differential equations)
  - not for individual runs, but average
- Discrete stochastic approach
  - discrete system evolution, via discrete events for reactions
  - obtain discrete-state stochastic process

# Discrete stochastic approach

- Work with states as vectors $\underline{x}$ of molecule counts for each species
  - probability $P(\underline{x},t)$ that at time t there will be $\underline{x}_A$ of species A
- The good news!
  - if constant state-dependent rates, obtain CTMC
  - therefore, can use stochastic process algebras as model description languages
- The stochastic approach admits
  - discrete event simulation
  - numerical solution (probabilistic model checking)
  - and is realistic for a single time course evolution, not just average

# Fragment of FGF pathway

- Fragment of Fibroblast Growth Factor (FGF) pathway
  - regulator of skeletal development, e.g. number of digits




- Biological challenges
  - unknown function of molecules, model different hypotheses
  - expensive experimental scenarios
- Aim to develop ODE and discrete stochastic models
  - ODE: use Cellarator & Mathematica
  - discrete: simulation (BioSPI, SPiM), verification (PRISM)

# FGF fragment – The reactions

1: FGF binds/releases FGFR

| | |
|---|---|
| FGF + FGFR → FGFR:FGF | k1=5e+8 M-1s-1 |
| FGF + FGFR ← FGFR:FGF | k2=0.002 s-1 |

2: Phosphorylation of FGFR (whilst FGFR:FGF)

| | |
|---|---|
| FGFR1 → FGFR1P | k3=0.1 s-1 |
| FGFR2 → FGFR2P | k4=0.1 s-1 |

3: Dephosphorylation of FGFR

| | |
|---|---|
| FGFR1P → FGFR1 | k5=0.1s-1 |
| FGFR2P → FGFR2 | k6=0.1s-1 |

4: Effectors bind phosphorylated FGFR

| | |
|---|---|
| SRC + FGFR1P → SRC:FGFR | k7=1e+6 M-1s-1 |
| SRC + FGFR1P ← SRC:FGFR | k8=0.02 s-1 |
| GRB2 + FGFR2P → GRB2:FGFR | k9=1e+6 M-1s-1 |
| GRB2 + FGFR2P ← GRB2:FGFR | k10=0.02 s-1 |

5: Relocation of FGFR (whilst SRC:FGFR)

| | |
|---|---|
| SRC:FGFR → relocFGFR | k11=1.1e-3 s-1 |

# FGF fragment – The modelling approach

- Consider a hypothesis about interaction between molecular species in the FGF pathway
  - obtain a set of ODEs from reactions, plot time trajectories for average concentrations (Cellerator)
  - model as a stochastic pi-calculus process, simulate to obtain individual time trajectories (BioSPI, SPiM)
  - model in reactive modules, analyse using probabilistic model checking (PRISM)

- Probabilistic model checking, as opposed to simulation
  - wide range of quantitative properties
  - compute for range of parameters: quantitative trends
  - can definitively establish causal relationships
  - able to identify best/worst case scenarios
  - but suffers from state explosion problems

# Stochastic π-calculus code fragment

```
FGFR ::= FGFR_FGF_0 | FGFR_Ph1_0 | ...

FGFR_FGF_0 ::= reloc1?[], true ;                  % relocation
   bind_fgf!{ rel_fgf, reloc4 }, FGFR_FGF_1.      % binding FGF
FGFR_FGF_1 ::=  rel_fgf?[] , FGFR_FGF_0;          % releasing FGF
            ph1?[] , FGFR_FGF_1;                  % phosphorylation
            reloc1?[] , reloc4 ! [] , true;       % relocation ...

FGFR_Ph1_0 ::= ph1![] , FGFR_Ph1_1 .              % phosphorylation
FGFR_Ph1_1 ::= dph1![] , FGFR_Ph1_1;              % dephosphorylation
   bind_src!{rel_src1, rel_src2 } , FGFR_SRC. % binding Src

FGFR_SRC ::= rel_src1?[], FGFR_Ph1_1 ;            % releasing Src
   dph1![], rel_src2![], FGFR_Ph1_0;              % dephos (& release Src)
   reloc![], reloc1![], reloc2![] , true.         % relocation
```

# Simple PRISM Example

1. $A+B \leftrightarrow A:B$ (binding/unbinding rates $r_1/r_2$)
2. $A \rightarrow$ (degradation rate $r_3$)

```
module A
  a : [0..1] init 1
  [bind] a=1 → r1 : (a'=0);
  [rel] a=0 → r1 : (a'=1);
  [] a=1 → r1 : (a'=0);
endmodule
```

```
module B
  b : [0..1] init 1
  [bind] b=1 → (b'=0);
  [rel] b=0 → (b'=1);
endmodule
```

```
module AB
  ab : [0..1] init 0
  [bind] ab=0 → (ab'=1);
  [rel] ab=1 → (ab'=0);
endmodule
```

reward structure 1:
time A and B are bound

```
rewards "r1"
  ab=1 : 1;
endrewards
```

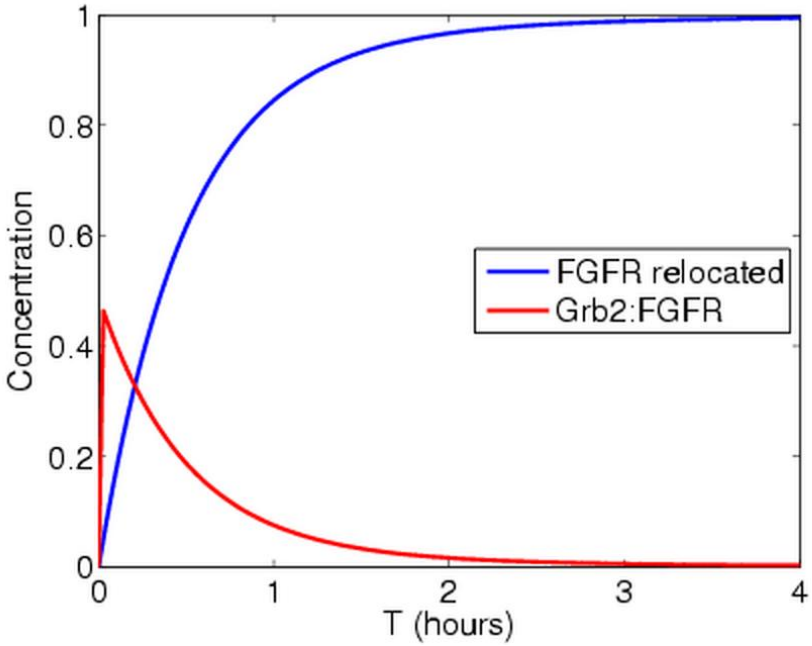
reward structure 2:
binding of A & B

```
rewards "r2"
  [bind] true : 1;
endrewards
```

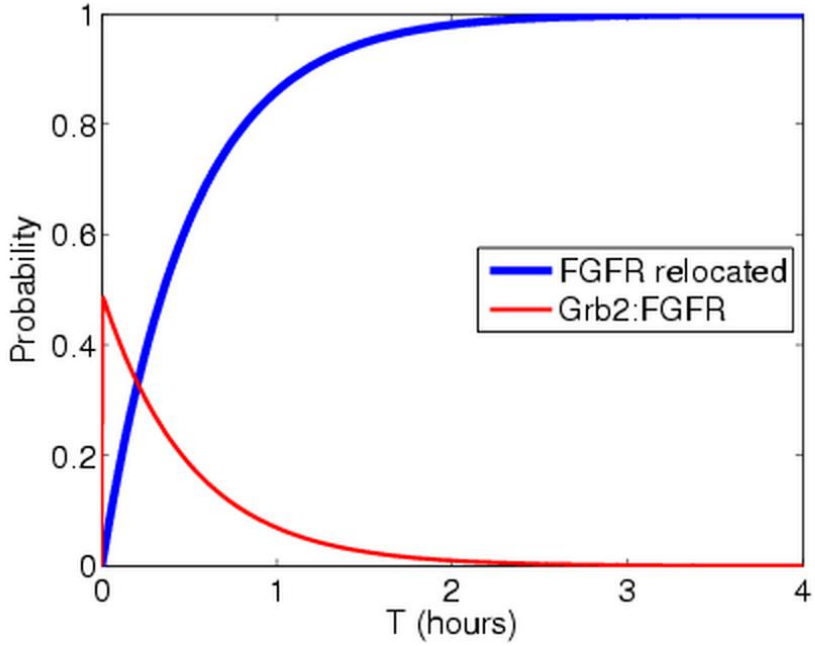
# FGF fragment – Results

Concentration/quantity of two forms of FGFR over time

ODEs

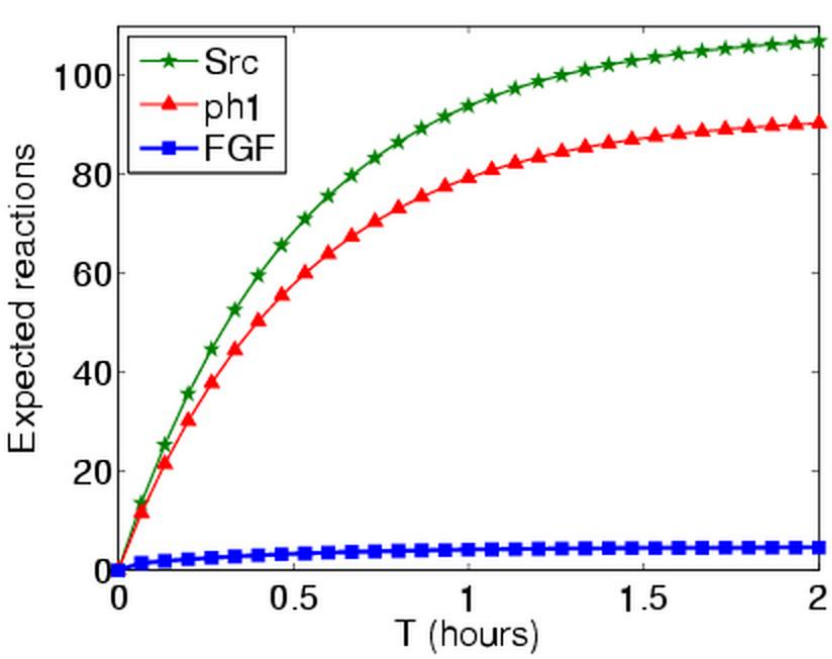



BioPEPA (50 runs)

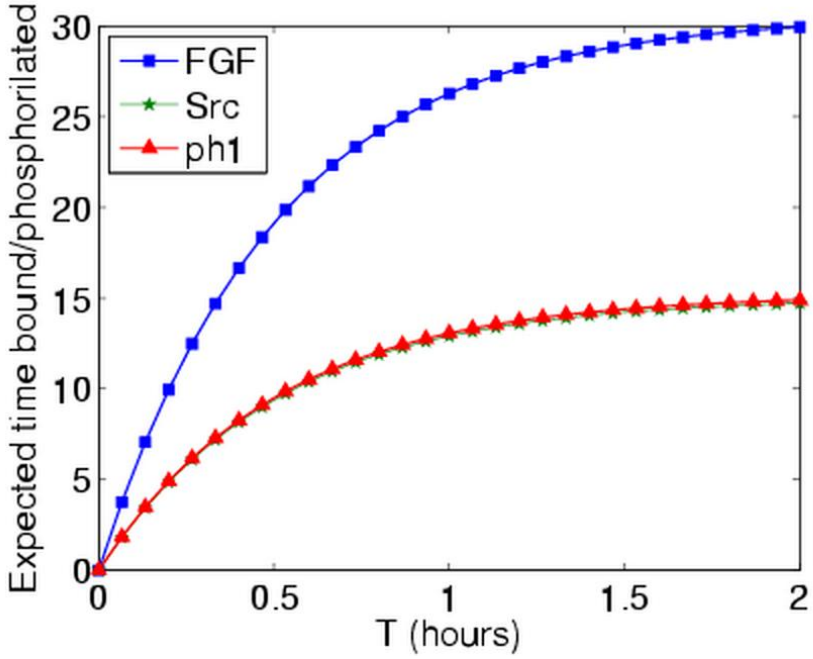


# FGF fragment – PRISM results $R_{=?}[C \leq T]$

Expected number of reactions by time T
(assign reward 1 to transitions in which the reaction occurs)




Expected time complex spends bound up to time T
(assign reward 1 to states in which the complex is bound)



# A variant of the FGF fragment

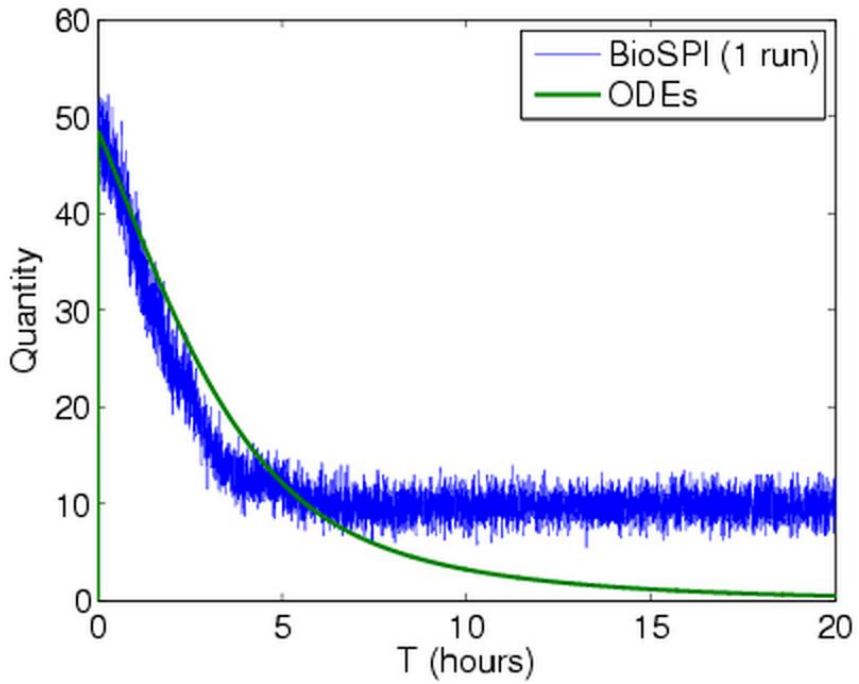
- Src positively regulates FGFR signalling by recruiting non-activated FGFR to the membrane, add reaction:

$$\text{FGFR:Src} \rightarrow \text{FGFR:Src} + \text{FGFR} + \text{Src}$$

Change initial amount of Src from 100 to 10 molecules, and similarly for ODEs

Difference between ODE and BioSPI caused by stochastic approach more accurate when number of molecules small

i.e. Src cannot be totally degraded in ODE



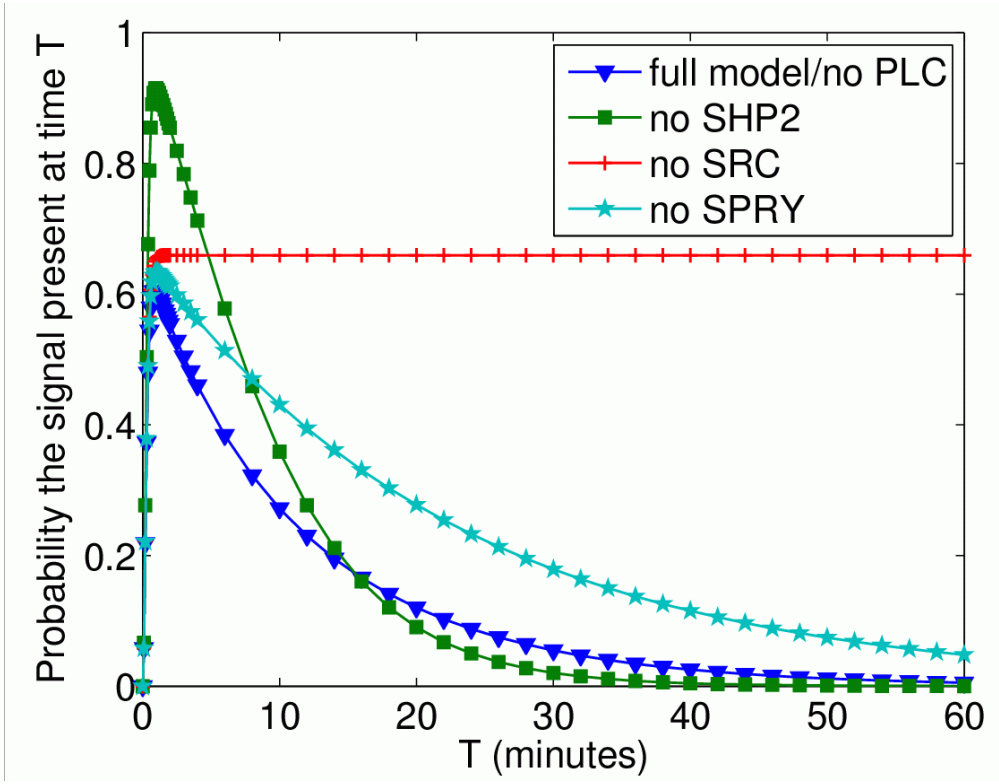
# PRISM model of full FGF pathway

- Biological Model
  - 12 elements
  - 14 phosphorylation sites
  - 14 sets of reaction rules (38 rules)

- PRISM model
  - one element of each type (10 modules and 26 variables)
  - relatively small state space
    (80,616 states and 560,520 transitions)
  - however, highly complex: large number of interactions
  - ODE model > 300 equations

# FGF pathway – Model checking results

- Probability Grb2 bound to FRS2 at time T
  - $P_{=?}$ [ true $U^{[T,T]}$ $a_{Grb2}$ ]



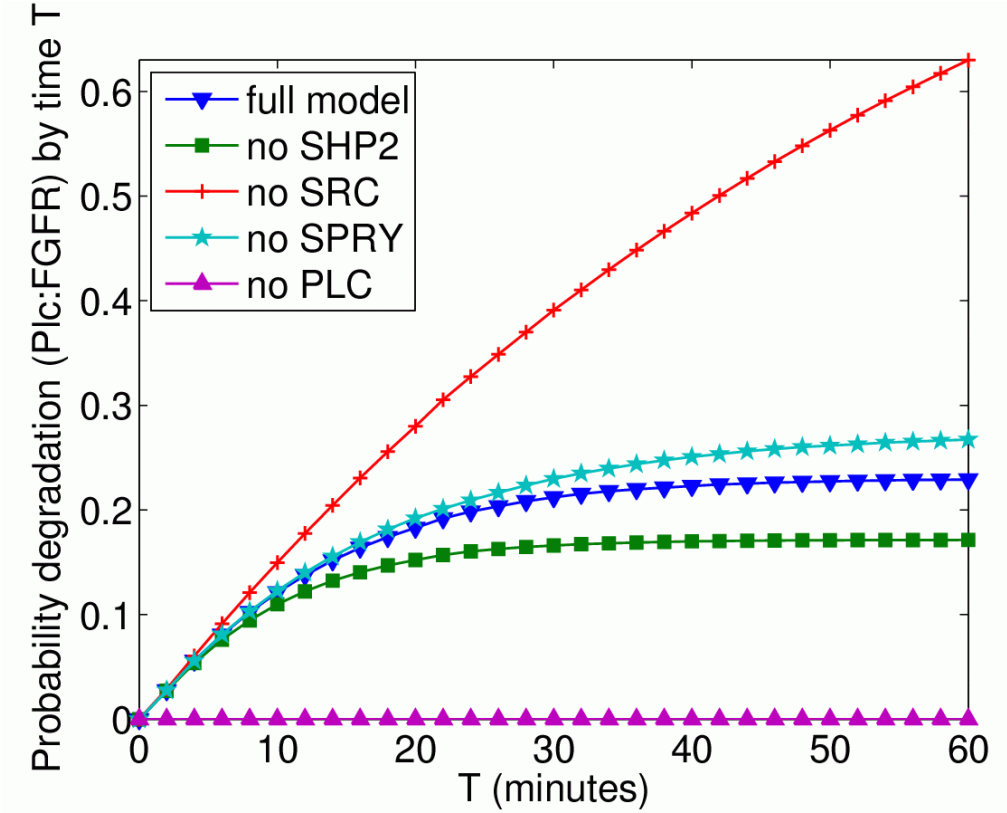

no SRC: no relocation of FRS2, and hence the signal can remain active

no SHP2: main cause of FRS2 dephosphorylation lost increasing the chance that:

- Grb2 bound to FRS
  faster increase in signal
- SRC bound to FRS2
  faster degradation in signal

# FGF pathway – Model checking results

- Probability PLC causes degradation/relocation by T
  - $P_{=?}$ [ $\neg(a_{src} \vee a_{spry} \vee a_{plc})$ $U^{[0,T]}$ $a_{plc}$ ]



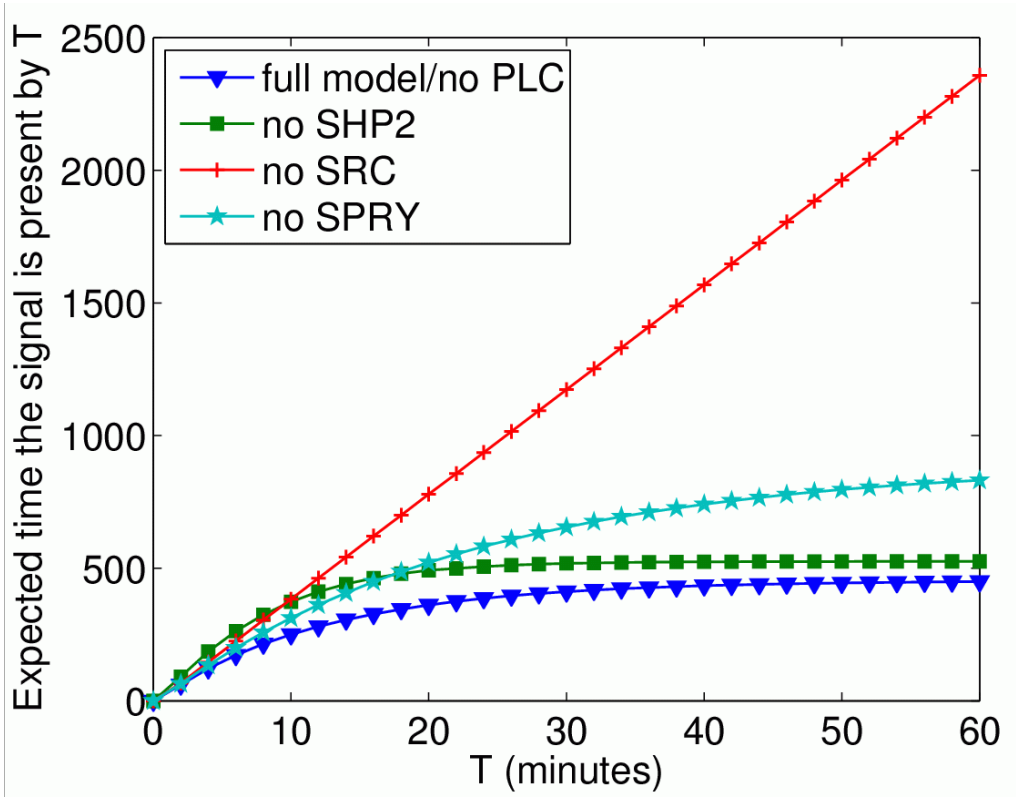

no PLC: PLC cannot cause degradation

no SRC: FRS2 not relocated, more chance of degradation by PLC

no SHP2: greater chance SRC bound to FRS2, increasing the possibility of FRS2 causing relocation

# FGF pathway – Model checking results

- Expected time GRB2 bound to FRS2 within time T
  - $R_{=?}$ [ C≤T ] (assign reward 1 to states where Grb2:FRS2)

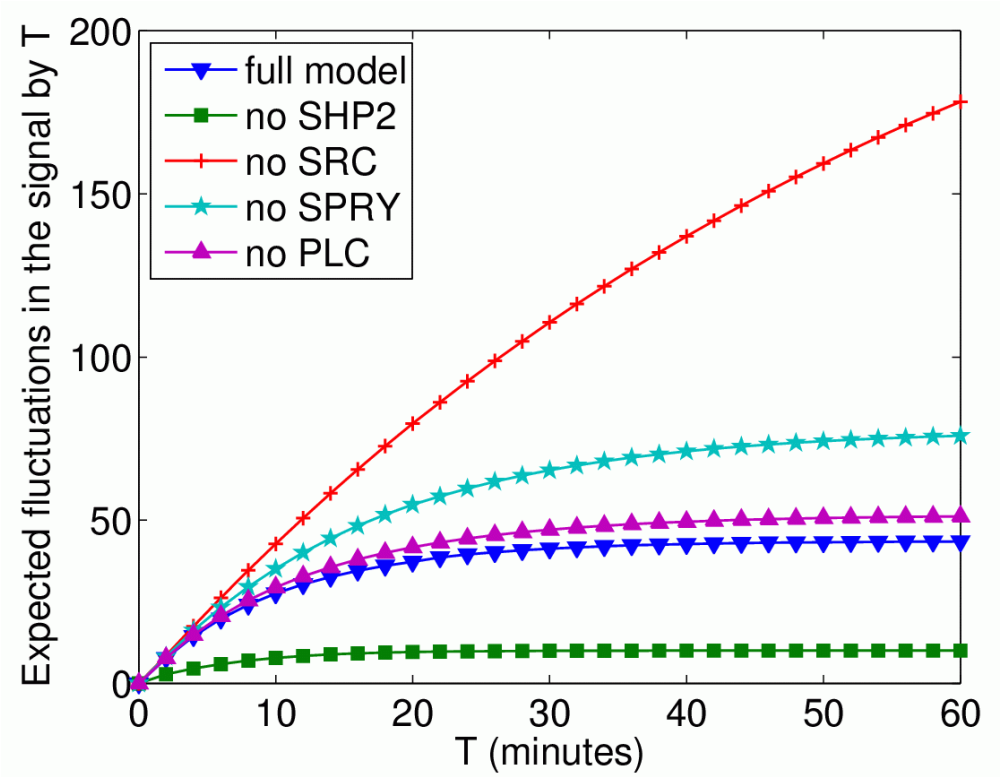



No SRC: no relocation of FRS2 and greater chance FRS2 remains active for longer, hence GRB2 and FRS2 spend more time bound

SPRY: no degradation of FRS2, again GRB2 and FRS2 spend more time bound (but SPRY has smaller influence than SRC)

# FGF pathway – Model checking results

- Expected number of times GRB2 & FRS2 bind by T
  - $R_{=?}$ [ C≤T ] (assign reward 1 to transitions binding Grb2/FRS2)




Cases when SRC and SPRY removed: increased chance that FRS2 remains active, and hence GRB2 and FRS2 can bind more often

No SHP2: decrease in the chance that GRB2:FRS2 unbind, therefore the chance that GRB2 and FRS2 are in a position to (re)bind decreases

# Summing up…

- What have we achieved?

- For dynamic power management
  - formulated a methodology for analysing power management policies
  - probability and expectation
  - constraints include buffer size, number of messages, etc
  - since applied by others, e.g. [SMA+07]

- For biological signalling
  - applied probabilistic model checking to test a range of detailed quantitative queries not possible with simulation
  - identified predictions, confirmed experimentally

# Further information

- More on the power management case study
  - see [NPK+05]
- More on FGF pathway
  - see [HKN+06]

- More on similar systems
  - power scavenging [SMA+07]
  - RKIP inhibited ERK pathway [CVGO06]
  - molecular systems [BCM+05]

- More information, see the PRISM web page

  www.prismmodelchecker.org