

Modelization and Simulation of Nano Devices in nano κ calculus

A. Credi¹, M. Garavelli¹, C. Laneve²,
S. Pradalier³, S. Silvi¹, and G. Zavattaro²

¹ Dipartimento di Chimica “G. Ciamician”, Università di Bologna

² Dipartimento di Scienze dell’Informazione, Università di Bologna

³ Ecole Polytechnique, Paris

Abstract. We develop a process calculus – the **nanok** calculus – for modeling, analyzing and predicting the properties of molecular devices. The **nanok** calculus is equipped with a simple stochastic model, that we use to model and simulate the behaviour of a molecular shuttle, a basic nano device currently used for building more complex systems.

1 Introduction

In 2006 the University of Bologna funded an interdisciplinary project of its Departments of Chemistry and Computer Science – the *CompReNDe Project* (Compositional and executable Representations of Nano Devices). The project combines the expertises of two groups, one specialized in the design and construction of devices and machines of molecular size [3, 2] and the other one qualified in formal models, based on the theory of process calculi, for describing and analyzing molecular systems [7, 14]. Such expertises are joined together in order to accomplish three main endeavours: (i) deliver a programming model for describing molecular machines that is also amenable to automated simulations and verifications by means of existing algorithms, (ii) apply the model for a formal analysis of real cases of molecular machines to possibly reveal complex behaviours that have not been experimentally observed yet, and (iii) use the simulations as tools to assist chemists in the design of novel nano devices.

The CompReNDe research activity started with the initial goal of formalizing a [2]rotaxane [20] into the κ calculus [7] in order to simulate its behaviour *in silico* by means of some contemporary stochastic evaluator [10, 19, 5]. [2]rotaxanes [20] (simply rotaxanes in the following) are systems composed of a molecular axle surrounded by a ring-type (macrocylic) molecule. Bulky chemical moieties (“stoppers”) are placed at the extremities of the axle to prevent the disassembly of the system. In rotaxanes containing two different recognition sites on the axle (“stations”), it is possible to switch the position of the macrocylic ring between the two stations by an external energy input as illustrated in Figure 1. Several rotaxanes of this kind, known as *molecular shuttles*, have been already developed (see [6] and the references therein) and used for building more complex systems [13, 12, 2].

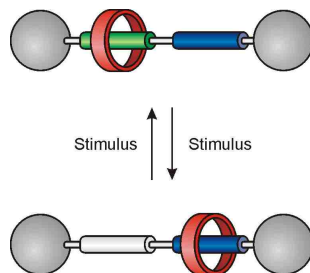


Fig. 1. Schematic representation of a two-station rotaxane and its operation as a controllable molecular shuttle.

The κ calculus is a formal language idealizing protein-protein interactions, as a particular restricted kind of graph-rewriting. Bindings are explicit: proteins are nodes with fixed numbers of sites, complexes are connected graphs built over such nodes where bonds are represented by *names*. Biological reactions are modeled by two kinds of rewriting rules: *complexations*, which create bonds, and *decomplexations*, which destroy bonds. Notably, the κ calculus has been compiled into π -calculus [16] introducing a finer-grained concurrent model, the $\mathfrak{m}\kappa$ calculus, where reactions have to be at most binary. The significant property of $\mathfrak{m}\kappa$ calculus is to be *protein-centred*, rather than reaction-centred as it was the case for κ calculus, thus being amenable to distributed implementations.

We therefore undertook the formalization of a molecular shuttle in $\mathfrak{m}\kappa$ calculus and we soon realized that such calculus was inadequate as well. The $\mathfrak{m}\kappa$ calculus is much too verbose because it compels designers to reason in terms of bonds and complexations and decomplexations. There are reactions that are neither complexations nor decomplexations, such as the *ion exchanges*. These reactions, used in our molecular shuttle to stimulate the movement of the macrocyclic ring, might be implemented by sequences of complexations and decomplexations, thus changing the granularity of the chemical semantics. The $\mathfrak{m}\kappa$ calculus model is much too abstract because it overlooks quantitative aspects. Such aspects, in particular reaction rates and the derived stochastic semantics are a must for providing meaningful simulations of molecular machines.

We overcome these inadequacies of the $\mathfrak{m}\kappa$ calculus by defining a new model, the **nanok** calculus, having three types of reactions – *creations*, *destructions*, and *exchanges* – and retaining a stochastic semantics. This stochastic semantics is problematic for the **nanok** calculus because it uses names for representing molecular bonds. In this respect, our model is close to Milner’s pi calculus [16]. However, instead of following the techniques of the stochastic pi calculus [18], we have preferred for **nanok** calculus to extend Cardelli’s language of stochastic interacting processes [4]. In facts, in this way, we get a simple model that may be easily simulated or verified by means of existing well known algorithms [9].

We then apply the **nanok** calculus to describe and analyze an instance of rotaxane, RaH [15, 1], for which the dynamic behaviour has been experimen-

tally characterized in detail [8]. We have considered two groups of simulations. The first ones are used to validate the model, checking whether the experiments reproduced *in silico* coincides with those already performed *in vitro*. The second ones simulate *in silico* the expected behaviour of the rotaxane RaH under conditions not yet observed *in vitro*. Interestingly, we show that under extreme conditions of very low concentration of rotaxane RaH, some of the assumptions, usually taken about the behaviour of the rotaxane in standard conditions of concentration, are no longer valid.

2 The nano κ calculus: syntax and semantics

Two disjoint countable sets of names will be used: a set of *species*, ranged over by A, B, C, \dots ; and a totally ordered set of *bonds*, ranged over by x, y, z, \dots . Species are sorted according to the number of *fields* and *sites* they possess. Let $\mathfrak{s}_f(\cdot)$ and $\mathfrak{s}_s(\cdot)$ be two functions returning naturals; the integers $1, 2, \dots, \mathfrak{s}_f(A)$ and $1, 2, \dots, \mathfrak{s}_s(A)$ are respectively the fields and the sites of A . ($\mathfrak{s}_f(A) = 0$ means there is no field; $\mathfrak{s}_s(A) = 0$ means there is no site). In the following, fields are ranged over by h, i, j, \dots ; sites are ranged over by a, b, c, \dots .

Sites may be either *bound* to other sites or *unbound*, i.e. not connected to other sites. The state of sites are defined by injective maps, called *interfaces* and ranged over by σ, ρ, \dots . Given a species A , its interfaces are partial functions from $\{1, \dots, \mathfrak{s}_s(A)\}$ to the set of bonds or a special empty value ε . A site a is bound with bond x in σ if $\sigma(a) = x$; it is unbound if $\sigma(a) = \varepsilon$. For instance, if A is a species with three sites, $(2 \mapsto x, 3 \mapsto \varepsilon)$ is one of its interfaces. In order to ease the reading, we write this map as $2^x + 3$ (the empty value is always omitted). This interface σ does not define the state of the site 1, which may be bound or not. In the following, when we write $\sigma + \sigma'$ we assume that the domains of σ and σ' are disjoint. Interfaces, being injective on bonds, cannot express that the endpoints of a bond belong to the same species (*cf. self complexation* in [7]). This design choice simplifies the presentation of nano κ calculus.

Fields represent the internal state of a species. The values of fields are defined by maps, called *evaluations*, and ranged over by u, v, \dots . For instance, if A is a species with three fields, $[1 \mapsto 5, 2 \mapsto 0, 3 \mapsto 4]$ is an evaluation of its. As before, we write this map as $1^5 + 2^0 + 3^4$. We assume there are finitely many internal states, that is every field h is mapped into values in $\{0, \dots, n_h\}$. In the following, we use partial evaluations and, when we write the union of evaluations $u + v$, we implicitly assume that the domains of u and v are disjoint.

Definition 1 (Molecules and Solutions). A molecule $A[u](\sigma)$ is a term where u is a total map on the fields of A . Solutions, ranged over by S, T, \dots , are defined by the following grammar

$$S ::= A[u](\sigma) \mid S, S$$

The operator “,” is assumed to be associative, so $(S, S'), S''$ is equal to $S, (S', S'')$ (and we always omit parentheses).

Solutions retain the property that bond names always occurs exactly twice. Let \emptyset be the empty map. We use the following shorthand notations: $A(\sigma)$ instead of $A[\emptyset](\sigma)$, $A[u]$ instead of $A[u](\emptyset)$, and simply A instead of $A\emptyset$.

Example 1. As a running example we consider two typical chemical reactions:

- $Na + Cl \longleftrightarrow Na^+ + Cl^-$ (sodium chloride) and
- $H + H \longleftrightarrow H_2$ (hydrogen gas) .

In the first reaction, an ion is exchanged between two instances of species Na and Cl . The molecules of the two species can be in two possible states: either they have the extra ion Na^+ and miss an ion Cl^- or they are in the states with all the ions Na and Cl . We model these two possible states using one field ion with values 0 and 1 respectively denoting the absence or the presence of the ion. Formally we can use $Na[ion^0]$ and $Na[ion^1]$ for Na and Na^+ , and $Cl[ion^0]$ and $Cl[ion^1]$ for Cl^- and Cl , respectively.

The second chemical reaction represents the creation/destruction of a bond between two hydrogen atoms. This may be described by using a site 1 and bond names. For instance, the solution with H_2 is modelled by $H(1^x), H(1^x)$. An unbound instance of hydrogen is simply represented by H , as its evaluation and interface are both empty.)

Definition 2 (Reactions). *Reactions of nanok calculus are either creations C, or destructions D, or exchanges E. The format of the first two types is $((A, a, u, u', \sigma), (B, b, v, v', \phi), \lambda)$; while the format of exchanges is $((A, u, u', \rho, \rho'), (B, v, v', \psi, \psi'), \lambda)$, such that:*

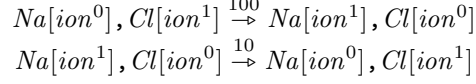
1. $\text{dom}(u') = \text{dom}(u)$ and u and u' are partial evaluations of A , $\text{dom}(v') = \text{dom}(v)$ and v and v' are partial evaluations of B ,
2. $\text{ran}(\sigma) = \text{ran}(\phi)$ and σ and ϕ are interfaces of A and B , respectively, such that $a \notin \text{dom}(\sigma)$ and $b \notin \text{dom}(\phi)$;
3. (for exchanges) ρ, ρ' and ψ, ψ' are interfaces of A and B , respectively, with $\text{ran}(\rho') = \text{ran}(\psi)$ and either $\rho = \rho'$ and $\psi = \psi'$ or $\rho = a^x + \rho'', \rho' = a + \rho''$ and $\psi = b + \psi'', \psi' = b^x + \psi''$;
4. and $\lambda \in \mathbb{R}^+ \cup \{\infty\}$.

For readability's sake, we write creations as $A[u](a+\sigma), B[v](b+\phi) \xrightarrow{\lambda} A[u'](a^x + \sigma), B[v'](b^x + \phi)$, destructions as $A[u](a^x + \sigma), B[v](b^x + \phi) \xrightarrow{\lambda} A[u'](a + \sigma), B[v'](b + \phi)$, and exchanges as $A[u](\rho), B[v](\psi) \xrightarrow{\lambda} A[u'](\rho'), B[v'](\psi')$.

The difference between the three kinds of rules is concerned with the modification of the interfaces: *creations* produce a new bond between the two unbound sites a and b , *destructions* remove the bond between the sites a and b , while *exchanges* either leave the interfaces unchanged or move one bond from a reactant to the other (bond-flipping exchange).¹

¹ The terms creation and destruction have been preferred to *complexation* and *decomplexation* used in [7, 14] because they have a more neutral chemical meaning.

Example 2. The **nanok** calculus reactions that corresponds to the two reactions of the sodium chloride are



where we have considered a rate 100 for the left to right direction and 10 for the right to left direction.

The **nanok** calculus reactions that corresponds to the two reactions of the hydrogen gas are



where the right direction has been given rate 5 and the left direction has been given rate 0.05.

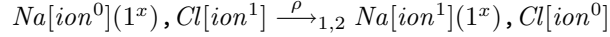
The formal definition of *reactants* and the corresponding *products* of reactions follows. We use μ to range over ρ_L, ι, x and ρ_R, ι, x and ρ_L, ι and ρ_R, ι and ρ . Let $\bar{\mu}$ be the following operation (notice that $\bar{\bar{\mu}} = \mu$):

$$\bar{\mu} \stackrel{\text{def}}{=} \begin{cases} \rho_R, \iota, x & \text{if } \mu = \rho_L, \iota, x \\ \rho_L, \iota, x & \text{if } \mu = \rho_R, \iota, x \\ \rho_R, \iota & \text{if } \mu = \rho_L, \iota \\ \rho_L, \iota & \text{if } \mu = \rho_R, \iota \\ \rho & \text{if } \mu = \rho \end{cases}$$

Definition 3 (Basic transition relation). *The basic transition relation of solutions, written $\xrightarrow{\mu}_{\ell} \cup \xrightarrow{\mu}_{\ell, \ell'}$, is the least relation that satisfies the following rules (ι are always injective renamings on bonds):*

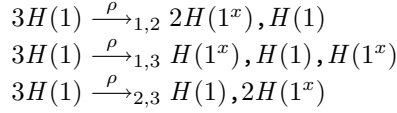
- (creations) if $\rho = A[u](a + \sigma), B[v](b + \phi) \xrightarrow{\lambda} A[u'](a^x + \sigma), B[v'](b^x + \phi)$ and $\text{dom}(\iota) = \text{ran}(\sigma) (= \text{ran}(\phi))$ and $z \notin \text{ran}(\sigma \circ \iota + \nu)$ then both $A[u + w](a + \sigma \circ \iota + \nu) \xrightarrow{\rho_L, \iota, z}_1 A[u' + w](a^z + \sigma \circ \iota + \nu)$ and $B[v + w](b + \phi \circ \iota + \nu) \xrightarrow{\rho_R, \iota, z}_1 B[v' + w](b^z + \phi \circ \iota + \nu)$;
- (destructions) if $\rho = A[u](a^x + \sigma), B[v](b^x + \phi) \xrightarrow{\lambda} A[u'](a + \sigma), B[v'](b + \phi)$ and $\text{dom}(\iota) = \text{ran}(\sigma) (= \text{ran}(\phi))$ then both $A[u + w](a^x + \sigma \circ \iota + \nu) \xrightarrow{\rho_L, \iota, x}_1 A[u' + w](a + \sigma \circ \iota + \nu)$ and $B[v + w](b^x + \phi \circ \iota + \nu) \xrightarrow{\rho_R, \iota, x}_1 B[v' + w](b + \phi \circ \iota + \nu)$;
- (exchanges) if $\rho = A[u](\sigma), B[v](\phi) \xrightarrow{\lambda} A[u'](\sigma'), B[v'](\phi')$ and $\text{dom}(\iota) = \text{ran}(\sigma) (= \text{ran}(\phi))$ then both $A[u + w](\sigma \circ \iota + \nu) \xrightarrow{\rho_L, \iota, v}_1 A[u' + w](\sigma' \circ \iota + \nu)$ and $B[v + w](\phi \circ \iota + \nu) \xrightarrow{\rho_R, \iota, v}_1 B[v' + w](\phi' \circ \iota + \nu)$, where v is either ε or $\iota(x)$, according to $\text{ran}(\sigma) \setminus \text{ran}(\sigma')$ is \emptyset or $\{x\}$;
- (lifts) if $S \xrightarrow{\mu}_{\ell} S'$ and, when ρ is a creation, $(\text{name}(S') \setminus \text{name}(S)) \cap \text{name}(T) = \emptyset$, then both $S, T \xrightarrow{\mu}_{\ell} S', T$ and $T, S \xrightarrow{\mu}_{\ell' + \ell} T, S'$, where T has ℓ' molecules;
- (communications) if $S \xrightarrow{\mu}_{\ell} S'$ and $T \xrightarrow{\bar{\mu}}_{\ell'} T'$ then $S, T \xrightarrow{\rho}_{\ell, \ell' + \ell'} S', T'$, where ρ is the rule of μ and S has ℓ'' molecules. If ρ is a creation, then the bond used by the reaction is the least one that is not used in S, T .

The basic transition relation definitely deserves to be spelled out. A reaction, such as $Na[ion^0], Cl[ion^1] \xrightarrow{100} Na[ion^1], Cl[ion^0]$ is a *schema*, namely it only addresses the fields and the sites of the reactants that are useful for the reaction. For example, it may be the case that Na retains a site to be used for other complexes, such as the *sodium peroxide*. In this case, the rule may be applied either to $Na[ion^0]$, where the site is unbound, or to $Na[ion^0](1^x)$. In this latter case, the reaction is instantiated as the transition:

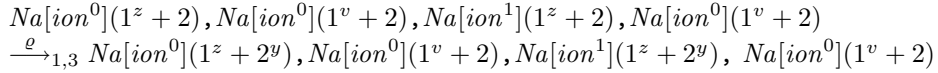


The basic transition relation is indexed by numbers. Since the solutions are sequences, these numbers give the exact positions of the reactants in the sequences. In the first three cases, the position is always 1 because the solution consists of one molecule. In the fourth case, the index is increased by the number of the molecules on the left, if any. The last case models a reaction: the solution is split into two parts S and T containing the reactants at positions ℓ and ℓ' , respectively. In the composite solution S, T , the reactants are at ℓ and $\ell'' + \ell'$, where ℓ'' is the number of molecules of S . For example let kM be $\underbrace{M, \dots, M}_{k \text{ times}}$ and

let ρ be the hydrogen gas reaction. The following transitions are possible



The basic transition relation is labelled by finite injective renamings. To clarify this point, consider the creation $\varrho = Na(1^x + 2), Na(1^x + 2) \xrightarrow{10} Na(1^x + 2^y), Na(1^x + 2^y)$ (a bond is created between two sodium molecules provided they are already bound). Then take the solution $Na[ion^0](1^z + 2), Na[ion^0](1^v + 2), Na[ion^1](1^z + 2), Na[ion^0](1^v + 2)$. We derive the expected transition



following a structured operational semantics approach [17]. Namely, we focus on the single reactants and lift the transitions to “,”-contextes. This is correct inasmuch as one records the instantiation of bonds in the left-hand sides of reactions with the actual names of the molecules: the two reactants must instantiate bonds in the same way. This is the reason why the first two molecules of the above solution cannot react with ϱ . More precisely, $Na[ion^0](1^z + 2) \xrightarrow{\varrho_L, \iota, y}_1 Na[ion^0](1^z + 2^y)$, where $\iota = [x \mapsto z]$, and $Na[ion^0](1^v + 2) \xrightarrow{\varrho_R, \iota, y}_1$.

Our final remarks regard the fourth and fifth items of Definition 3. Whenever $S \xrightarrow{\rho, \iota, x}_\ell T$ and ρ is a creation, the basic transition relation also admits $S \xrightarrow{\rho, \iota, y}_\ell T\{y/x\}$, where y is fresh. This nondeterminism is removed when the reaction occurs because the bond has to be the least name not occurring in S . It is also worth to notice that there is no rule lifting a transition $\xrightarrow{\mu}_{\ell, \ell'}$ to a context “,”:

we use the associativity of \equiv , to partition a solution S into S', S'' such that the reactants are in S' and S'' .

The basic transition relation is excessively intensional to be sensible for chemistry. Consider a solution containing hundreds of molecules of the species A and B that could react with ρ . The relation $\xrightarrow{\mu}_{\ell, \ell'}$ distinguishes the two pairs of reactants, and this is not possible in practice. More reasonably, the transition relation should represent *collectively* all the possible combinations of one molecule of species A with one molecule of species B . For instance, the solution A, A, B transits with $\xrightarrow{\rho}_{1,3}$ and $\xrightarrow{\rho}_{2,3}$. Abstracting out the order of the molecules, we obtain a unique transition whose rate is twice the rate of ρ . However quotienting the solutions with commutativity axioms of “ \equiv ,” does not yield an adequate extensionality. In fact, when ρ is a destruction, between A and B , the solution $A(a^x), A(a^y), B(a^x), B(a^y)$ transits with $\xrightarrow{\rho}_{1,3}$ and $\xrightarrow{\rho}_{2,4}$ into two solutions that cannot be equated by permutations of the molecules in the solution. In these cases one has to use injective renamings of bonds.

Definition 4 (Structural equivalence). *The structural equivalence between solutions, noted \equiv , is the least equivalence satisfying the following two rules (we remind that solutions are already quotiented by associativity of “ \equiv ”):*

1. $S, T \equiv T, S$;
2. $S \equiv T$ if there exists an injective renaming ι on bonds such that $S = \iota(T)$.

Example 3. Commutativity and injective renaming of the structural equivalence permit to prove the two following equivalences, respectively

$$Na[h^0], Cl[h^1] \equiv Cl[h^1], Na[h^0] \quad H(b^x), H(b^x) \equiv H(b^y), H(b^y)$$

Combining both commutativity and injective renaming we can prove that

$$H(b^x), H(b^x), H(b^z), H(b^z) \equiv H(b^y), H(b^k), H(b^k), H(b^y)$$

Proposition 1. *Let $S \equiv S'$.*

1. If $S \xrightarrow{\mu}_{\ell} T$ then there is T' and a renaming ι such that $S' \xrightarrow{\iota(\mu)}_{\ell'} T'$ and $T' \equiv S'$;
2. if $S \xrightarrow{\rho}_{\ell, \ell'} T$ then there is T' such that $S' \xrightarrow{\rho}_{\ell', \ell''} T'$ and $T' \equiv S'$.

The following notations are relevant for the definition of the stochastic transition relation:

- $rate(\rho)$ returns the rate of the reaction ρ ;
- $next(S) = \{(\rho_{\ell, \ell'}, T) \mid S \xrightarrow{\rho}_{\ell, \ell'} T\}$; $next_{\infty}(S) = \{(\rho_{\ell, \ell'}, T) \mid S \xrightarrow{\rho}_{\ell, \ell'} T \text{ and } rate(\rho) = \infty\}$;
- S has *finite rates* if, for every $(\rho_{\ell, \ell'}, T) \in \mathcal{S}$, $rate(\rho)$ is not ∞ ;
- let \mathcal{S} be a set of pairs (X, T') (the second element is a solution; the first one is not specified), $[\mathcal{S}]_T$ is the subset of \mathcal{S} of those pairs (X, T') such that $T' \equiv T$;

- $\text{can}(\mathcal{S})$ is defined over sets of pairs (X, T) such that the solutions occurring as second element of the pairs are all structurally equivalent. It returns a solution S such that there is X with $(X, \mathcal{S}) \in \mathsf{S}$.

Definition 5 (Stochastic transition relation). *The **nanok** calculus stochastic transition relation \mapsto^λ , where $\lambda \in \mathbb{R}^+ \cup \{\infty\}$, is the least relation satisfying the following rules:*

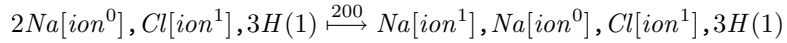
- if $\mathsf{S} \xrightarrow{\rho}_{\ell, \ell'} \mathsf{T}$ and $\text{rate}(\rho) = \infty$ then $\mathsf{S} \mapsto^\infty \text{can}([\text{next}_\infty(\mathsf{S})]_{\mathsf{T}})$;
- if $\mathsf{S} \xrightarrow{\rho}_{\ell, \ell'} \mathsf{T}$ and $\text{next}(\mathsf{S})$ has finite rates then $\mathsf{S} \mapsto^\lambda \text{can}([\text{next}(\mathsf{S})]_{\mathsf{T}})$, where

$$\lambda = \sum_{(\rho_{\ell, \ell'}, \mathsf{T}') \in [\text{next}(\mathsf{S})]_{\mathsf{T}}} \text{rate}(\rho)$$

We notice that, by definition, the **nanok** calculus stochastic transition system is such that there is no state with outgoing \mapsto^∞ and \mapsto^λ (λ finite) transitions. Hereafter, the states with \mapsto^∞ outgoing transitions are called *transient states*, the other ones are called *markovian states*.

The interrelation between basic and stochastic transition relations is as follows: the stochastic one partitions the products of a solution (according to the basic transition relation) into equivalence classes, takes a canonical representative of the class, and defines a transition whose label is the sum of the rates of the reactions in the basic one that yield solutions in the equivalence class.

Example 4. As examples of stochastic transitions, we consider the reactions of sodium chloride (called ρ) and hydrogen gas (called ρ') of Example 1 for the solution $2\text{Na}[\text{ion}^0], \text{Cl}[\text{ion}^1], 3\text{H}$. This solution may transit with $\xrightarrow{\rho}_{1,3}$ and $\xrightarrow{\rho}_{2,3}$ into solutions that are structural equivalent. Therefore we obtain a unique stochastic transition:



We also observe that there is a unique transition $\xrightarrow{15}$ outgoing the initial solution and corresponding to $\xrightarrow{\rho'}_{4,5}$, $\xrightarrow{\rho'}_{4,6}$, and $\xrightarrow{\rho'}_{5,6}$.

3 Markov Chains and the **nanok** calculus

The stochastic transition relation of **nanok** calculus corresponds to an *Interactive Markov Chain (IMC) transition system* with only *silent interactive transitions* [11]. These transitions, which are those labelled ∞ in our model, are executed in the IMC model instantaneously and under the *maximal progress assumption*. That is, the so-called *sojourn time* in a transient state is 0, which amounts to favour silent interactive transitions to those labelled with finite rates (called *markovian* transitions). On the contrary, in a markovian state with n outgoing markovian transitions labelled $\lambda_1, \dots, \lambda_n$, the probability that the

sojourn time is less than t is exponentially distributed with rate $\sum_i \lambda_i$, i.e. $\text{Prob}\{\text{delay} < t\} = 1 - e^{-t \sum_i \lambda_i}$, and the probability that the j -th transition is taken is $\lambda_j / (\sum_i \lambda_i)$.

However the models underlying traditional simulation algorithms such as [9] are *Continuous Time Markov transition systems (CTMC)* that do not include interactive transitions. Having a CTMC is therefore primary to run automatic analysis tools for experimenting *in silico* the dynamics of nano-machines specifications in `nanok` calculus.

The mismatch between IMC with only silent actions and CTMC systems is due to two main reasons: (i) the nondeterminism and (ii) the persistency of the silent interactive transitions. As regards (i), consider two silent actions that apply to the same reactants and give two different products. If these products have only markovian transitions it is not possible to collect them in a unique solution. As regards (ii), if an infinite sequence of silent interactive transitions exists then the simulation time of the CTMC system will not advance anymore. Therefore collapsing all these transitions, by identifying the initial and final solutions of the sequence, is again not possible.

However, there are cases where the downgrading of an IMC system to a CTMC one is possible without modifying the semantics. This is *when all silent actions may be partitioned into confluent directed acyclic graphs of finite depth*. In fact, when the silent interactive transitions are partitioned into confluent directed acyclic graphs, there are no loops (there is no infinite sequence of silent interactive transitions), and all sequences of silent interactive transitions starting from the same state share the same final state, to which the initial state may be safely collapsed. The meaning of this collapse is that we are removing a finite amount of work which is performed in zero time.

The formal definition of downgrading of IMC to CTMC systems follows. We first introduce the auxiliary function *next markovian state* defined on solutions and yielding sets:

$$- \text{nextm}(\mathcal{S}) = \{((\lambda, \mathcal{T}'), \mathcal{T}) \mid \mathcal{S} \xrightarrow{\lambda} \mathcal{T}' \xrightarrow{\infty}^* \mathcal{T} \text{ and } \lambda \in \mathbb{R}^+ \text{ and } \mathcal{T} \not\xrightarrow{\infty}\}$$

We notice that $\text{nextm}(\mathcal{S})$ is undefined when \mathcal{S} is transient.

Definition 6 (Downgrading of an IMC system). *An IMC system $(\mathcal{S}, \xrightarrow{\lambda})$ is strictly-markovian if*

1. *states are either transient or markovian and*
2. *every subsystem consisting of silent interactive transitions is a confluent direct acyclic graph of finite depth.*

Let $(\mathcal{S}, \xrightarrow{\lambda})$ be strictly-markovian; the transition relation $\xrightarrow{\nu}$, where $\nu \in \mathbb{R}^+$, is the least one such that:

- *if \mathcal{S} is markovian then $\mathcal{S} \xrightarrow{\nu} \text{can}([\text{nextm}(\mathcal{S})]_{\mathcal{T}})$ with*

$$\nu = \sum_{((\lambda, \mathcal{T}'), \mathcal{T}'') \in [\text{nextm}(\mathcal{S})]_{\mathcal{T}}} \lambda$$

It is easy to verify that the relation $\stackrel{\nu}{\rightleftharpoons}$ defines a CTMC system. Moreover, the properties below are direct consequences of the construction:

- the probability distribution of the sojourn time in a markovian state is the same in the IMC and in the downgraded CTMC and
- the probability that one of the paths $S \xrightarrow{\lambda} \xrightarrow{\infty} T'$ with $T' \equiv T$ is taken in the IMC corresponds to the probability the unique transition $S \xrightarrow{\lambda'} T''$, with $T'' \equiv T$, is taken in the downgraded CTMC.

Actually the correspondence between strictly-markovian IMC and the associated CTMC is much stronger: the IMC semantics, the *markovian bisimulation* [11], is still a markovian bisimulation on the CTMC when restricted to its states. We will detail the correspondence in the full paper.

4 nano κ calculus at work: the rotaxane case study

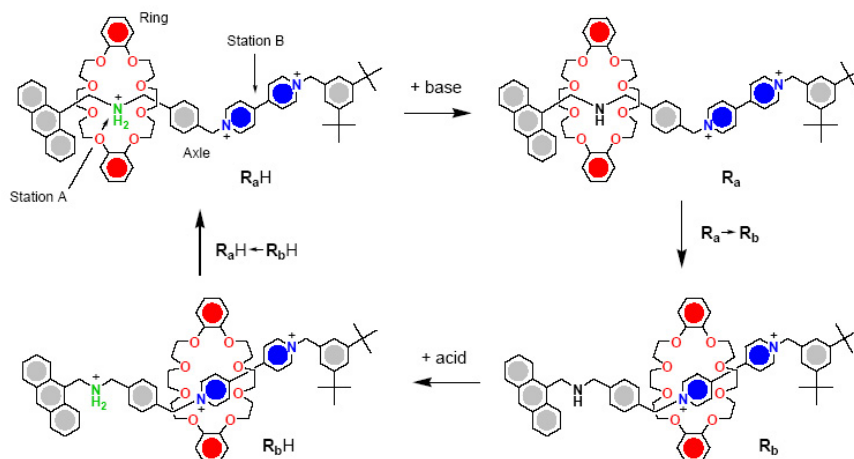


Fig. 2. Schematic representation of the shuttling processes of the molecular ring in the examined rotaxane.

The investigated rotaxane RaH (Figure 2) [15, 1] is made of a stoppered axle containing an ammonium (A) and an electron acceptor bipyridinium (B) stations that can establish hydrogen-bonding and charge-transfer interactions, respectively, with the ring component, which is a crown ether with electron donor properties. Since the hydrogen bonding interactions between the macrocyclic ring and the ammonium center are much stronger than the charge-transfer interactions of the ring with the bipyridinium unit, the rotaxane exists as only one of the two possible translational isomers, denoted as RaH in Figure 2. In

solution, addition of a base (e.g., tributylamine) converts the ammonium center into an amine function, giving the transient state Ra that is transformed into the stable state Rb as a consequence of the displacement of the macrocycle onto the B station. The process can be reversed by addition of acid (e.g., trifluoroacetic acid) and the initial state is restored, passing through the transient state denoted as RbH. Nuclear magnetic resonance, absorption and luminescence spectroscopic experiments, together with electrochemical measurements, indicate that the acid-base controlled switching, which is fully reversible and relatively fast, exhibits a clear-cut on-off behaviour [1].

The Rotaxane RaH is particularly appropriate to test the modeling approach described in the present paper because it is one of the very few cases wherein not only the thermodynamic properties, but also the dynamic behavior of the system have been experimentally characterized in detail. Specifically, the macrocycle's shuttling process between the ammonium/amine and bipyridinium stations in this rotaxane, driven by the successive addition of base and acid, have been investigated in solution [8]. The rate constants for the "forward" (Ra→Rb) and "backward" (RbH→RaH) shuttling motions (vertical processes in Figure 2) of the molecular ring, which occur, respectively, upon deprotonation and reprotonation of the ammonium/amine recognition site on the axle (horizontal processes in Figure 2), were found to be $0.72s^{-1}$ and $40s^{-1}$ at $293^{\circ}K$, respectively.

4.1 Modeling the rotaxane RaH in nanoκ calculus

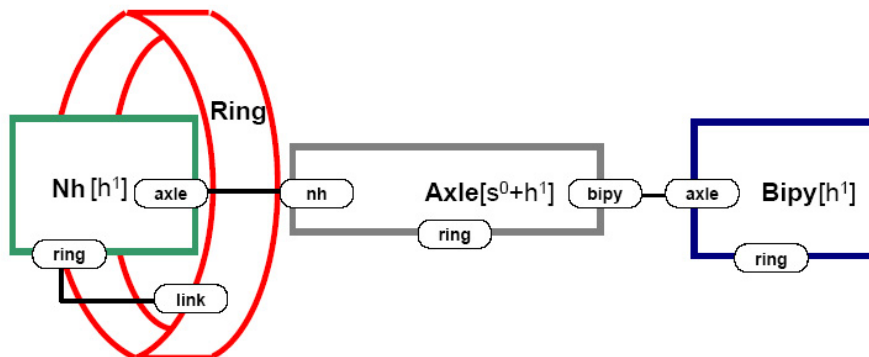


Fig. 3. Initial state of the Rotaxane RaH in nanoκ calculus.

The nanoκ calculus molecules. Figure 3 illustrates the nanoκ calculus modeling of the rotaxane RaH. We use four species:

- *Nh* models the ammonium/amine station of the rotaxane: it has one field h and two sites *ring* and *axle*;

- *Axle* models the spacer between the two stations: it has two fields s and h and three sites nh , $bipy$, and $ring$;
- *Bipy* models the bipyridinium station: it has one field h and two sites $ring$ and $axle$;
- *Ring* models the crown ether ring: it has no field and one site $link$.

The pairs of sites $axle$ of Nh and nh of $Axle$, and $axle$ of $Bipy$ and $bipy$ of $Axle$ are always linked in our modeling. They model the covalent bonds maintaining the structural integrity of the axle. Exactly one site $ring$ of Nh , $Bipy$, and $Axle$ is linked at a given moment at $link$ of $Ring$. The first two cases respectively model the “stable” RaH and Rb states of Figure 2 in which the ring is steadily located around the Nh or the $Bipy$ molecules, respectively. The last case models the “unstable” states; these are the Ra and RbH states of Figure 2 in which the ring is not steadily located. In order to distinguish between the Ra and RbH states, we use the field s of the $Axle$: it holds the value 0 if the ring is around the Nh (Ra state), 1 if it is around the $Bipy$ (RbH state).

Ammonium and amine functions have different chemical nature but can be seen as protonated and deprotonated version of the same species. Thus we model both by the same \mathbf{nanok} calculus species Nh . Its field h is used to record the presence or absence of a proton on Nh : its value is 1 if it is protonated, and 0 otherwise.

As $Ring$'s movements are triggered by protonations and deprotonations due to acid-base reactions, we also need to have acid and base molecules in our modeling. We consider the species $Acid$ and $Base$ both with one field h having value 1 in case the acid/base molecule holds the proton to be exchanged, 0 otherwise (for instance $Acid[h^1]$ and $Base[h^0]$ are respectively an acid molecule ready to give a proton and a base molecule ready to receive a proton).

The initial state for rotaxane RaH is thus modeled by the term:

$$Nh[h^1](axle^s + ring^x) , Axle[s^0 + h^1](nh^s + bipy^r + ring) , \\ Bipy[h^1](axle^r + ring) , Ring(link^x)$$

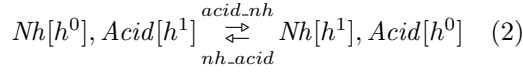
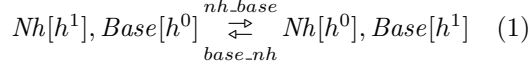
graphically depicted in Figure 3.

Note that the Nh is initially protonated (and this information is present also in the $Axle$ and the $Bipy$), the $Axle$ is bound to the Nh and the $Bipy$, and the $Ring$ is bound to the Nh .

The \mathbf{nanok} calculus reactions. We now present the reactions used in our modeling. Reactions 1, 2, 7 and 8 are presented with a double arrow (that are *reversible* reactions). Formally they correspond to two \mathbf{nanok} calculus reactions, one achieved reading the reaction from left to right considering the rate over the arrow, and another one achieved reading it from right to left considering the rate below. In this section we do not consider numerical values of rates, this is detailed in part 4.2.

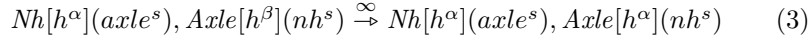
A base can get the proton of a protonated Nh , and a Nh can get a proton from an acid. These acid-base reactions are reversible. Reactions 1 and 2 model

this phenomena. The systems corresponding to the left-hand side and right-hand side coexist, even if one can be much predominant according to the ratio $nh_base/base_nh$ (and $acid_nh/nh_acid$).

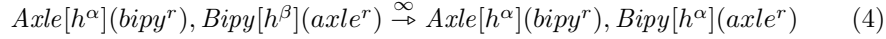


The protonation state of the molecule Nh needs to be known by $Bipy$ because it affects its interaction with $Ring$. Reactions 3 and 4 achieve this by passing information from Nh to $Bipy$ through $Axle$. These updates are instantaneous because the reactions have infinite rates (this is relevant for the correctness of our simulation, since these reactions have no counterpart in chemistry).

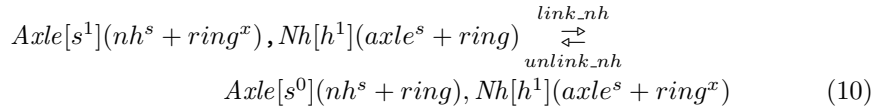
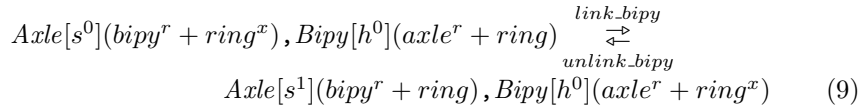
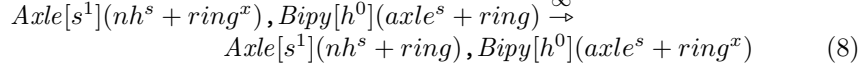
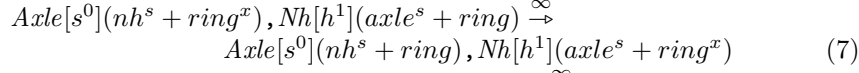
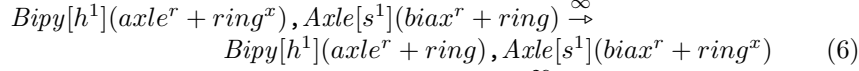
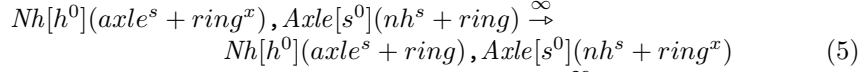
if ($\alpha \neq \beta$)



and:



We achieve the modeling of $Ring$ movements in two steps. Firstly the instantaneous reactions to deprotonation/reprotonation (reactions 5–8), and secondly the actual $Ring$ shuttling (reactions 9 and 10). As reactions 5–8 represent immediate consequences of deprotonation or reprotonation of Nh , they have infinite rates. Reactions 9 and 10 are reversible, because the $Ring$ is susceptible to return to the previous station due to the Brownian motion.



4.2 Simulation results

It is not difficult to verify that the above modeling of rotaxane RaH in \mathbf{nanok} calculus yields a strictly markovian IMC system. Therefore we may safely down-

grade it to a CTMC system that we use to simulate *in silico* the behaviour of the rotaxane RaH.

As previously discussed the rates for the ring movements are respectively $link_bipy = 0.72$ and $link_nh = 40$. On the basis of the estimated equilibrium constants, the rates for the reverse reactions are quantified two orders of magnitude smaller, i.e. $unlink_bipy = 0.0072$ and $unlink_nh = 0.4$.

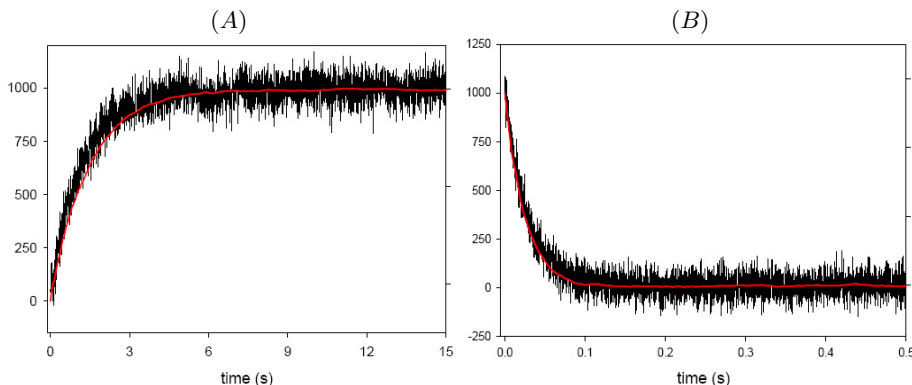


Fig. 4. Comparing the simulations *in silico* with the experiments *in vitro*. Grey traces: number of *Rings* located around *Bipys* during the “forward” Ra→Rb (part A) and the “backward” RbH→RaH (part B). Black traces: UV absorbance changes observed upon the occurrence of the same respective shuttling processes.

The aim of the first two simulations depicted in Figure 4 is to check whether the experimentation *in silico* can reproduce the results observed in *in vitro* [8]. The techniques used for the *in vitro* experimentation did not permit to observe and quantify the deprotonation/reprotonation rates (this is not surprising as these are very fast acid-base reactions). Thus, in the simulation we have considered instantaneous deprotonation/reprotonation, i.e. $nh_base = acid_nh = \infty$ and $base_nh = nh_acid = 0$. In both simulations, we have considered 1000 rotaxanes: in the first one we have simulated deprotonation and “forward” (Ra→Rb) shuttling, in the second one reprotonation and “backward” (RbH→RaH) shuttling. In the first simulation the shuttling phase is completed in around 6 seconds, while in the second one in 0.1 seconds; this is a consequence of the different rates of the two directions of shuttling. Very remarkably, simulated data are in strike agreement with the experimental results.

After these initial encouraging results, we have decided to use the *in silico* simulation techniques to provide a comprehensive view of the overall reactions depicted in Figure 2, simulating also the deprotonation/reprotonation phases not observed in the *in silico* experimentation. More precisely, the aim of this second group of simulations was to either validate or invalidate the assumption according to which deprotonation/reprotonation can be considered “instanta-

neous” with respect to the shuttling time. To this aim, we have simulated deprotonation/reprotonation under two different concentrations of rotaxanes. In fact, this is a bimolecular reaction whose rate is influenced by the concentration of the reactants. For instance, at the concentration considered in [8], i.e. $10^{-5}M$, assuming 1000 instances of rotaxane and base/acid, a plausible rate for deprotonation/reprotonation is $2 \times 10^2 s^{-1}$ (with reverse reaction rate on the order of $2 \times 10^{-5} s^{-1}$) while at the concentration $10^{-8}M$ it is $0.2 s^{-1}$ (with reverse reaction on the order of $0.2 \times 10^{-7} s^{-1}$).

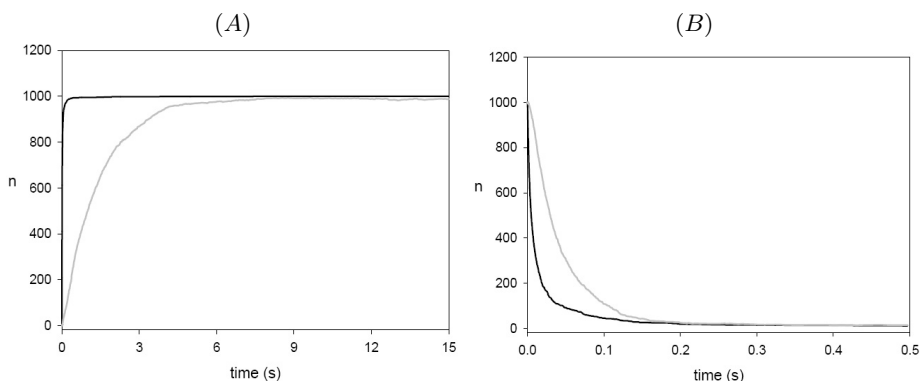


Fig. 5. Number of *Rings* located around *Bipys* (grey trace) and number of protonated rotaxanes (black trace) during the “forward” shuttling in the presence of base molecules (part A) and the “backward” shuttling in the presence of acid molecules (part B).

We have performed the two simulations, namely deprotonation with subsequent “forward” shuttling and reprotonation with subsequent “backward” shuttling, considering the two different concentrations.

The results at concentration $10^{-5}M$ are not reported in the paper as they essentially confirm the validity of the “instantaneous” deprotonation/reprotonation assumption. We report in Figure 5 the results for concentration $10^{-8}M$ as they are definitely more interesting; the rings start moving before the deprotonation/reprotonation phase is over. This proves that in the rotaxane RaH the stimulus and the subsequent shuttling could interplay, and this opens interesting scenarios that requires further investigation. For instance, it could be the case that using weak acid/base molecules (for which the ratio between the deprotonation/reprotonation rate and the reverse rate is smaller) the interplay between the stimulus and the shuttling could give rise to currently unknown emerging behaviours.

References

1. P. R. Ashton, R. Ballardini, V. Balzani, A. Credi I. Baxter, M. C. T. Fyfe, M. T. Gandolfi, M. Gómez-López, M.-V. Martínez-Díaz, A. Piersanti, N. Spencer, J. F.

- Stoddart, M. Venturi, A. J. P. White, and D. J. Williams. Acid-base controllable molecular shuttles. *Journal of Am. Chem. Soc.*, 120:11932–11942, 1998.
2. J. D. Badjic, V. Balzani, A. Credi, S. Silvi, and J. F. Stoddart. A molecular elevator. *Science*, 303:1845–1849, 2004.
 3. Vincenzo Balzani, Alberto Credi, and Margherita Venturi. Molecular devices and machines – concepts and perspectives for the nano world. In *Wiley-VCH*, Weinheim, 2007.
 4. Luca Cardelli. On process rate semantics. 2007.
 5. Luca Cardelli and Andrew Phillips. Spim homepage. At research.microsoft.com/~aphillip/spim/, 2006.
 6. Benoit Champin, Pierre Mobian, and Jean-Pierre Sauvage. Transition metal complexes as molecular machine prototypes. *Chemical Society Reviews*, 36(2):358–366, 2006.
 7. Vincent Danos and Cosimo Laneve. Formal molecular biology. *Theoretical Computer Science*, 325(1):69–110, 2004.
 8. S. Garaudée, S. Silvi, M. Venturi, A. Credi, A.H. Flood, and J. F. Stoddart. Shuttling dynamics in an acid-base-switchable [2]rotaxane. *ChemPhysChem*, 6:2145, 2005.
 9. Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem*, 81:2340–2361, 1977.
 10. Stephen Gilmore and Jane Hillston. The pepa workbench: a tool to support a process algebra-based approach to performance modelling. In *Proc. of the 7th int. conference on Computer performance evaluation : modelling techniques and tools*, pages 353–368, Secaucus, NJ, USA, 1994. Springer-Verlag New York, Inc.
 11. H. Hermanns. *Interactive Markov Chains: The quest for quantitative quality*, volume 2428 of *LNCS*. Springer-Verlag, Amsterdam, 2002.
 12. T.-J. Huang, B. Brough, C.-M. Ho, Y. Liu, A.H. Flood, P.A. Bonvallet, H.-R. Tseng, J.F. Stoddart, M. Baller, and S. Magonov. A nanomechanical device based on linear molecular motors. *Applied Physics Letters*, 85(22):5391–5393, 2004.
 13. M.C. Jimenez, C. Dietrich-Buchecker, and J.-P. Sauvage. Towards synthetic molecular muscles: Contraction and stretching of a linear rotaxane dimer. *Angew. Chem. Int. Ed.*, 39(18):3284–3287, 2000.
 14. Cosimo Laneve and Fabien Tarissan. A simple calculus of protein and cells. In *Proceeding of MeCBIC'06*, volume 1677 of *ENTCS*, 2007.
 15. M.-V. Martínez-Díaz, N. Spencer, and J. F. Stoddart. The self-assembly of a switchable [2]rotaxane. *Angew. Chem. Int. Ed. Engl.*, 36:1904–1907, 1997.
 16. Robin Milner. *Communicating and mobile systems: the π -calculus*. Cambridge University Press, Cambridge, 1999.
 17. G. D. Plotkin. A Structural Approach to Operational Semantics. Technical Report DAIMI FN-19, University of Aarhus, 1981.
 18. Corrado Priami. Stochastic pi-calculus. *Computer Journal*, 38(7):578–589, 1995.
 19. Corrado Priami, Aviv Regev, Ehud Shapiro, and William Silverman. Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Information Processing Letters*, 80:25–31, 2001.
 20. J.-P. Sauvage and C. O. Dietrich-Buchecker (eds.). Molecular catenanes, rotaxanes and knots. In *Wiley-VCH*, Weinheim, 1999.