Fluidization of Discrete Event Models or a marriage between the discrete and the continuous

Manuel Silva
Universidad de Zaragoza

1. As an appetizer
... on models, systems and fluidization...

2. (Discrete) Petri Nets: elements to be used

3. Fluidization of autonomous discrete models

4. Fluid timed models: server semantics...

Informal, trying to provide intuition
Using examples as simple as possible

Well known, model is:
* in Science & Technology: the representation
* in Fine Arts: the reference...
... just the contrary.

The model?

How representative is a model? --- Fidelity vs complexity?

This system is discrete or continuous?

1) Continuous?
2) Discrete?
3) ???

The system & the model:
1) Purpose
2) Fidelity
3) Complexity

Models are “views”

... but if model is a representation, it is only “an image”, “a view”
constructed with a purpose, not “the reality”:
A MODEL vs THE (REAL) SYSTEM
The state explosion problem... (a very naive view...)

Homotecy

Increasing the population (or initial marking):
• the approximation improves
• much bigger computational savings

Markovian T-time interpretation

Steady State

The bigger the marking, the better the approximation & the computational savings

If we fluidify (from integer to reals) the model & use infinite server semantics (min operator)...
1. As an appetizer
2. (Discrete) Petri Nets: elements to be used
   - A quick global perspective of place/transition nets (P/T)
   - Fundamental equation and mathematical programming
   - Components, invariants & stable predicates
   - Removing spurious solutions & implicit places
   - Rank theorems: a proof of synergy between the performance and logic analysis
3. Fluidization of autonomous discrete models
4. Fluid timed models: server semantics...

Place/Transition nets (P/T, PNs)
- PNs as directed bipartite multigraph
- PNs and incidence matrices
- Places: local state variables
- Transitions: local state transformers
- Dynamic system
- Marking: numerical distributed state & evolution
- Synchronizations (two mechanisms...):
  - Joins / Rendez-Vous
  - Weights on arcs
- Fundamental or state (transition) equation
  - Let $C = Post - Pre$ (incidence matrix)
  - If $m$ is reachable (exist $e \in E$)
  - then $m = m_0 + C \sigma$
  - $(\sigma)$ is the firing count component of $t$

Fundamental equation & fluid relaxation
- With $m = m_0 + C \sigma$, two main computational problems:
  1) in the naturals (computational cost) &
  2) with spurious (validity of computations...)
- Our focus is today on:
  1) Fluidization: relaxation into the non-negative reals
  2) How to remove spurious solutions on:
     2.1) The discrete model: untimed & timed
     2.2) The fluid model: untimed & timed

... in the PN paradigm
Decomposed views... (I): Invariants

From the fundamental equation (structure):
1. Annuler vectors of the incidence matrix, C:
   * Left & non-negative: P-semiflows (γ):
     \[ \gamma \mathbf{C} = 0, \gamma \geq 0 \]
   * Right & non-negative: T-semiflows (ξ):
     \[ \mathbf{C} \mathbf{x} = 0, x \geq 0 \]
2. Invariants (laws...):
   * \[ \gamma \mathbf{m}_0 = \gamma \mathbf{m} \] --- token conservation law (P-inv.)
   * Firing & marking repetitive sequences law (T-inv.)
3. Subnets:
   * Conservative component (P-subnet)
   * Repetitive component (T-subnet)

Decomposed views... (II): stable predicates

Looking the net as a directed & bipartite graph (structure):
1. Inclusions concerning subsets of places (here we do not consider weights on arcs...):
   * Trap: subset of places s.t. the subset of output transitions are included in their input subset of transitions
   * Siphon: the inverse of a trap
2. Stable predicates (laws...):
   * Trap: if marked, cannot be fully unmarked (tokens are trapped)
   * Siphon: if emptied, cannot be marked again
3. Subnets:
   * Trap component (P-subnet)
   * Siphon component (P-subnet)

Decomposed views... (II): stable predicates

- \( p_1 \) is a trap & a siphon, but not a conservative component

Conservative components are trap & siphon components
- \( \gamma = (1 \ 0 \ 1) \) is a P-semiflow, thus: \( m_1 + m_2 = 10 \)
- \( \gamma = (0 \ 1 \ 1) \) is a P-semiflow, thus: \( m_1 + m_2 = 11 \)

How to cut the spurious deadlock \( m = (0 \ 1 \ 10) \)?

Removing a spurious solution that unmark a trap is a polynomial time problem

- \( p_1 \) is an initially marked trap, thus: \( m_1 \geq 1 \)
- \( m_1 + m_2 = 10 \), thus \( m_2 \leq 9 \)
- \( m_1 + m_2 = 11 \), thus \( m_2 \geq 2 \)
  \( (p_2 \) has 2 frozen tokens)

\( q_3 \) makes \( p_2 \) implicit
1. As an appetizer
2. (Discrete) Petri Nets: elements to be used
3. Fluidization of autonomous discrete models
   3.1 Fluidization of DES & fluidizability
   3.2. Basic properties of autonomous continuous models
   3.3. Improving the approximation (I): removing spurious solutions
4. Fluid timed models: server semantics...

3.1 Fluidization of DES: concept & fluidizability
   Definition: A continuous transition
   \[ \text{enab}(t, m) = \min \left\{ \frac{m[p]}{\text{Pre}[p, t]} \mid \text{Pre}[p, t] > 0 \right\} \]
   \[ m \xrightarrow{\alpha, t} m' \]
   \[ \alpha \in \mathbb{R}^+ \quad \alpha \leq \text{enab}(t, m) \]
   Neighbour places become continuous

Reachability (under total continuity)

* \( m \) is reachable from \( m_0 \) iff a finite sequence \( \sigma = \alpha_1 t_1 \cdots \alpha_k t_k \)
exists, such that \( m_0 \xrightarrow{\alpha_1 t_1} \cdots \xrightarrow{\alpha_k t_k} m \)
* Reachability space, \( \text{RS}(N, m_0) = \) set of reachable markings
  \[ \text{RS}_D(N, m_0) \subseteq \text{RS}_C(N, m_0) \]

Same reachability space

A main question:
When a given property is preserved by fluidization?
Fluidization: can be approximated by the scaling of the initial marking
- homothetic markings!
- a monotonicity property

Monotonicities
Marking monotonicity of property \( \Pi \)
\[ \Pi \text{ holds in } (N, m_0) \Rightarrow \]
\[ \Pi \text{ holds in } (N, m'_0), \forall m'_0 \geq m_0 \]

Marking homothetic monotonicity of property \( \Pi \)
\[ \Pi \text{ holds in } (N, m_0) \Rightarrow \Pi \text{ holds in } (N, k \cdot m_0), \forall k \in \mathbb{N}^+ \]

Remark: Monotonicity \( \Rightarrow \) Homothetic monotonicity

All (discrete) PN models can be fluidified?
Deadlock-freeness: discrete vs continuous (!!!)

Live: yes or not?

New concept: lim-reachability
- At the limit: a trap is emptied
- So traps are no more traps...!

Homothetic monotonous,
...but not monotonous

Deadlock-free
for the marking, and

Bigger marking
and deadlock!
3.2. Basic properties of autonomous continuous models

Let us assume in the sequel that:
1) The net $N$ is consistent ($\exists x > 0, Cx = 0$) and
2) all the transitions can be fired “a little” at $m_0$ 
(otherwise stated: no siphon is empty at $m_0$).

Thus $RS(N, m_0) = \{m | m = m_0 + C\sigma, \sigma \geq 0\}$

* * Spurious solutions?:
• NOT on the continuous model;
• But spurious solutions in the originally discrete net model are here integrated.
> This raises the question of how to remove it!

Marking homotety and Deadlock-freeness

| Discrete: DF for any $k$-$m_0$
| Continuous: DF, not lDF |

$\langle N, m_0 \rangle_D$ is homothetic $\iff \langle N, m_0 \rangle_C$ is DF

$\langle N, m_0 \rangle_D$ is $\text{lim}$-$\langle N, m_0 \rangle_C$ is $\text{lim}$-DF

Sufficient condition for DF...

Approaching DF

$2^k = 8 LS$

$1 LS + 4 \text{ var.} + 2 \text{ equ.}$

Firing languages are identical if the same $m_0$ and $tp$ & $tq$ are "silent" transitions.

No system has a solution, thus they are deadlock-free systems.

$1 LS$

$8 LS$

$\alpha$, $\beta$

$\alpha$, $\beta$

$\alpha$, $\beta$

$\alpha$, $\beta$

$\alpha$, $\beta$

$\alpha$, $\beta$

$\alpha$, $\beta$
3.3. Improving the autonomous (thus, in general, all timed) approximations: cutting spurious solutions

Complementary (efficient) techniques:
- Spurious solutions that empty a trap (like in discrete models; reachable in the limit)
- Spurious solutions that empty a siphon (feasible with finite sequences)
  if integer ... beyond Gomory cutting planes
- Enforce the bound to the bigger integer of Truncating
  the marking of places (addition of complementary places)
  the enabling of transitions (addition of self-loop places)
  ... below Gomory cutting planes

Decreasing improvement with increasing markings...

If the initial marking is “not too big”, any SB(p) & SF(p) should be truncated to natural values: floor functions

\[ SB(p3) = 1.5; \quad m(p3)+m(q3) = 1 \]

Adding \( q_3 \) avoids siphon \( S = (p_1, p_4) \) to be emptied in the continuous approximation

Adding \( q_4 \) does not prevent the non integer (deadlock) solution:
\[ m = (0 \ 0 \ 0 \ 1.5 \ 1.5) \]
Thus the integer hull is not obtained (too expensive for the benefits...)

4. Timed continuous models: semantics, properties and improvements

\[ m = m_0 + C \sigma \] (autonomous: state transition eq.)
... if the firing is a function of time
\[ m(\tau) = m_0 + C \sigma(\tau) \]
... deriving
\[ \dot{m}(\tau) = C f(\tau) \]
\[ m(0) = m_0 \] with \( f(\tau) = \sigma(\tau) \)

- How is \( f \) defined? \( \rightarrow \) firing semantics? Exist several...
- An interesting source of inspiration is Markovian PNs (inheritance of results, coherence with discrete performance models): lead to infinite server semantics

Infinite server and product (population) semantics

In the sequel we just concentrate on infinite server semantics
From configuration to operation mode

From the **graph** (structure):
1. **Configuration (allocation):** A set of p-t arcs defining a 1-cover of all transitions; those that constraint the firing...
2. **Region:** the sub-reachability space defined by a configuration (are disjoint except on the boundary)
3. **Operation mode:** the dynamic linear system corresponding to a configuration

**1. Configuration matrix:** \( \Pi = [p_r(p)] \)
- \( \Pi(p, p') = \frac{p_r(p')}{p_r(p)} \) if \( p, p' \notin C_i \)
- \( \Pi(p, p') = 0 \) otherwise

**2. Operation mode:**
- \( \Pi(m) = \Pi, \text{ where } \Pi \text{ is the configuration matrix associated to } R_i \)
- \( m(\tau) = C \cdot f(\tau) = C \cdot A \cdot \Pi(m) \cdot m(\tau) \)
- \( \Pi(m) = \Pi \)

---

### Infinite servers semantics. Properties

- **Positive systems:** \( f(\tau) \geq 0, m(\tau) \geq 0 \)
- **Marking & time homotheties (duality):**
  - If \( m'(0) = k \cdot m(0) \) then \( m'(\tau) = k \cdot m(\tau) \)
  - \( f'(\tau) = k \cdot f(\tau) \)
- **But no superposition**, what lead to many counter-intuitive behaviors

---

### Counterintuitive properties (I)

- The **steady-state** of the throughput of the continuous PN relaxation may **not** be an upper bound of corresponding discrete models

**Discrete:** 0.801 f/t.u.  
**Continuous:** 0.535 f/t.u.  
**To fluidify does not improve!**
Counterintuitive properties (II)

- No performance monotonicities with respect to resources

\[
\begin{align*}
\text{M}(p_5) = 3 & \Rightarrow \chi (t_3) = 1.071 \\
\text{M}(p_5) = 4 & \Rightarrow \chi (t_3) = 0.535 \\
\text{M}(p_5) = 5 & \Rightarrow \text{Deadlock}
\end{align*}
\]

Counterintuitive properties (II)

- No performance monotonicities with respect to resources

\[
\begin{align*}
\text{M}(p_5) = 3 & \Rightarrow \chi (t_3) = 1.071 \\
\text{M}(p_5) = 4 & \Rightarrow \chi (t_3) = 0.535 \\
\text{M}(p_5) = 5 & \Rightarrow \text{Deadlock}
\end{align*}
\]

Counterintuitive properties (III): steady-state

- No performance monotonicities with respect to firing speeds
- Discontinuity / Bifurcation (loss of hyperbolicity)

Performance monotonicity

- Mono-T-semiflow reducible (M-T-S-red) nets are conservative with a single T-semiflow if Homothetic Equal Conflicts \([\text{Pre} (t_3) = q \text{ Pre} (t_3)]\) are reduced.

Let \(N\) be a M-T-S-red net. Assume that steady-states are reached for given firing speeds and initial markings. If the st-st configurations contains the support of a P-semiflow,

- \(< N, \lambda_1, m_1 > & \leq < N, \lambda_2, m_2 >\) with \(m_1 \leq m_2\), or
- \(< N, \lambda_1, m_1 > & \leq < N, \lambda_2, m_2 >\) with \(\lambda_1 \leq \lambda_2\)

then \(P = Q (\text{i.e., system 2 never slower!})\)

- Corollary: SL&B EQ net systems are throughput monotonous wrt the marking and the firing speeds.

M-T-S net

- Monotous operation mode

- Non-monotous operation mode

3 of 4 conf.

- \(m(p_1) + m(p_2) = 10\)
- \(m(p_1) + m(p_2) = 10\), only if \(m(p_1) = 5\)
- No P-component “covered by the configuration”
4.2. Patterns of behaviours

(1) Equilibrium (point attractor)

Patterns of behaviours

2) Basic oscillatory (linear)

Patterns

3) Oscillatory (non-linear) & timing depending collapse: Siphons

Flows and phase-portraits: collapse if $\lambda_2 > 10$

Two discontinuities wrt $\lambda_2$ (at 1.0 & 10.0)

High & low frequencies and collapse

+++ Small variations in parameters:
- Orbits (like in Lotka-Volterra predator-prey model),
- Limit cycles
- ...
4.3 Improving the timed fluidization

With a preliminary character: use all improvements for untimed models that remove spurious solutions!

In particular, those that:
1) empty a marked trap
2) empty a siphon (and are known to be spurious)
3) add complementary places ($q_i$) that truncate the structural bound of places
Also: 4) add self-loop places that truncate the enabling bound of transitions

Reuse...: Interleaving qualitative & quantitative

Removing the spurious deadlock $m=(0 1 10)^T$

Adding marking and flow truncating places

Adding marking and flow truncating places

4.3.1 Bound reaching problem & $\rho$-semantics

- What to do when $m_0$ is “not that big”, and exist transitions with enabling degree = 1 (particularly if a synchronization appears as weight on the arc from $p$ to $t$)?
- One possibility: to maintain the transition $t_1$ as discrete (problem similar to response-time in RC circuits).
- Alternative IDEA: To divide the load of tokens in such a way that the firing of the transition start with a threshold value: $k=\rho$
  - Question: a “good” value for $\rho$? (heuristic?)

The enabling degree of $t_1$

Infinite server semantics  Hybrid net  $\rho$-semantics

With an heuristic technique to choose the value for $\rho$
4.3.2 Stochastic approximation

- $m^0$ “very big”: applicable a kind of functional law of numbers or deterministic fluid limit
- $m^0$ “not that big”: applicable a kind of functional central limit theorem (in the line of Donsker theorem)

If the discrete PN model to be approximated is provided with exponential timing (Markovian PN, analyzable by means of a Markov Chain that is isomorphous to the reachability graph), then:

The fluid PN approximation lead to a system of stochastic piecewise affine differencial equations:

\[
\sum_{i} \lambda^i \Pi(m^i) m^i \Delta \tau = \text{diag}[f \Delta \tau]
\]

where \(\text{diag}[f \Delta \tau]\) is a diagonal matrix

\[
m_{k+1} = m_k + C \lambda \Pi(m_k) m_k \Delta \tau + C \nu_k
\]

where \(w = \lambda \Pi(m_k) m_k \Delta \tau + \nu_k = f \Delta \tau + v_k\)

STCPNs: particularly interesting when “moving” among two or more configurations

<table>
<thead>
<tr>
<th>$\lambda^i$</th>
<th>$\Pi$</th>
<th>$m^i$</th>
<th>$\lambda^i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>54.62</td>
<td>55</td>
<td>0.7%</td>
</tr>
<tr>
<td>1.5</td>
<td>53.87</td>
<td>55</td>
<td>2.1%</td>
</tr>
<tr>
<td>1.2</td>
<td>51.10</td>
<td>55</td>
<td>7.9%</td>
</tr>
<tr>
<td>1</td>
<td>46.65</td>
<td>55</td>
<td>17.9%</td>
</tr>
<tr>
<td>1.05</td>
<td>40.72</td>
<td>55</td>
<td>35.6%</td>
</tr>
<tr>
<td>1</td>
<td>29.97</td>
<td>55</td>
<td>83.5%</td>
</tr>
</tbody>
</table>

STCPNs: particularly interesting when “moving” among two or more configurations

Improvement when the discrete system in the steady-state switch among different regions

**Concluding remarks**

- $m = m^0 + C \lambda^i$ is:
  1) in the naturals (computational cost) &
  2) with spurious (care with computations ...)

Our focus today, was on:

1) Fluidization: relaxation into the non-negative reals
2) How to remove spurious solutions on:

The discrete and corresponding fluid models

(untimed & timed)

**GOAL:** To overcome the state explosion problem... in certain cases (the constraint: the existence of non-fluidizable PN models)

- The bigger the marking: the better the approximation &
  the bigger the saving on computations.
- Useful with untimed & timed net models
- Reuse / adaptation of results from discrete PNs theory.

**Integration in the PN paradigm** (economy, coherence, synergy)

- Partial fluid relaxations: (a kind of) Hybrid Petri Nets
Concluding remarks

The use of "good" timed fluid (technically hybrid) models:
- Parametric optimization (where to go?, dim. of buffers?)
- Sensibility analysis
- Observers
- Controllers (min. time; ...)
- Diagnosis under faulty behaviours
- ...

"Good" is related to:
- the relative absence of spurious solutions
- the level of fluidization: Hybrid vs continuous nets?
- In timed models: an appropriate server semantics (deterministic or stochastic)?

Structural
- Generic
- Punctual

Thank for your attention

Manuel Silva
Universidad de Zaragoza