# RECENT DEVELOPMENTS FOR MONO- OR MULTI-RATE PARALLEL REAL-TIME CO-SIMULATION

# EXTRAPOLATION AND SCHEDULING FOR MULTICORE ARCHITECTURES

Abir El Feki, Salah Eddine Saidi, Nicolas Pernet, Laurent Duval, Mongi Ben Gaid (IFPEN)

Yves Sorel, Daniel Simon (INRIA)

Cyril Faure (CEA)



# OUTLINE

# Background

- Co-simulation: context & challenges
- Real-time Co-Simulation: an open problem
- Improving parallelism with the RCosim approach: Refined Co-simulation
- Ensuring co-simulation accuracy with CHOPtrey extrapolation approach
- Mapping real-time constraints for HiL
- Future work



#### BACKGROUND

- Co-simulation: Alternative to monolithic simulation 

  Simulation of a complex system using several coupled subsystems
  - A subsystem is modeled using the most appropriate tool instead of using a single modeling software
  - Subsystems are modeled and run in a segregated manner → The equations of each model are integrated using a solver separately
  - During the execution models exchange data → A synchronization mechanism is used between the models, in such a way that models update their inputs and outputs according to assigned communication steps
  - Easy upgrade, reuse, and exchange of models





#### BACKGROUND

- Co-simulation: Alternative to monolithic simulation 

  Simulation of a complex system using several coupled subsystems
  - A subsystem is modeled using the most appropriate tool instead of using a single modeling software
  - Subsystems are modeled and run in a segregated manner → The equations of each model are integrated using a solver separately
  - During the execution models exchange data → A synchronization mechanism is used between the models, in such a way that models update their inputs and outputs according to assigned communication steps
  - Easy upgrade, reuse, and exchange of models
  - Heterogeneous ODE models → Time consuming simulations





Model 2

Model 1

Model 3





Model 4 Complex model  $\rightarrow$  Time consuming simulation





# BACKGROUND (CONT'D)

#### A multi-core co-simulation kernel: Why?

- System-level simulation leads to agglomerate models which are classically disconnected, increasing the CPU demand at simulation time
- Simulation time becomes more and more a metric for model complexity
- Most 0D/1D simulation tools have mono-core kernel while mono-core computers are endangered

# How long will this CPU power remain unused?







# OUTLINE

# Background

- Co-simulation: context & challenges
- Real-time Co-Simulation: an open problem
- Improving parallelism with the RCosim approach: Refined Co-simulation
- Ensuring co-simulation accuracy with CHOPtrey extrapolation approach
- Mapping real-time constraints for HiL
- Future work



SUSTAINABLE MOBILITY

#### REAL-TIME SIMULATION NEEDS FOR CPS VALIDATION





7 © 2017 IFPEN

SUSTAINABLE MOBILITY

#### REAL-TIME SIMULATION NEEDS FOR CPS VALIDATION



#### Hardware-in-the-Loop → real-time constraints



8 © 2017 IFPEN

Digicosme GT OVSTR - 26/04/2017

#### REAL-TIME SIMULATION NEEDS FOR CPS VALIDATION



Hardware-in-the-Loop → real-time constraints



Digicosme GT OVSTR - 26/04/2017

#### REAL-TIME SIMULATION NEEDS FOR CPS VALIDATION



Hardware-in-the-Loop → real-time constraints



# OUTLINE

# Background

- Co-simulation: context & challenges
- Real-time Co-Simulation: an open problem
- Improving parallelism with the RCosim approach: Refined Co-simulation
- Ensuring co-simulation accuracy with CHOPtrey extrapolation approach
- Mapping real-time constraints for HiL
- Summary and outlook



#### RCOSIM: REFINED CO-SIMULATION DATAFLOW GRAPH OF FMUS

- Inter FMU dependencies specified by the user
- Identify locally if Y is dependent on U or not
  - FMI gives relationships between each Y and U
  - With FMI each I/O is computed with a different FMU functions
- Build refined dependency graph
  - Vertices: operations, a set of FMU functions
    - updateOut, updateIn, and updateState
  - Directed edges: precedencies between operations
  - Ordinary Differential Equations (ODEs)
    - No algebraic loops
    - Directed Acyclic Graph (DAG)
- Apply a multi-core scheduling heuristic on the dataflow graph







# MUO-RCOSIM EXTEND RCOSIM TO HANDLE MULTI-RATE CO-SIMULATION





# MUO-RCOSIM EXTEND RCOSIM TO HANDLE MULTI-RATE CO-SIMULATION



- Multi-rate co-simulation
  - Update the I/O of each FMU according to its communication step
- Need for a repeatable pattern of the multi-rate graph execution
- Repeat each operation r<sub>i</sub>= HS/H(o<sub>i</sub>) times, HS=lcm (H(o<sub>1</sub>), H(o<sub>2</sub>), ..., H(o<sub>n</sub>))

• E.g:  $H_B = 2 \times H_A$ 





14

# MUO-RCOSIM EXTEND RCOSIM TO HANDLE MULTI-RATE CO-SIMULATION



- Multi-rate co-simulation
  - Update the I/O of each FMU according to its communication step
- Need for a repeatable pattern of the multi-rate graph execution
- Repeat each operation r<sub>i</sub>= HS/H(o<sub>i</sub>) times, HS=lcm (H(o<sub>1</sub>), H(o<sub>2</sub>), ..., H(o<sub>n</sub>))

• E.g: 
$$H_B = 2 \times H_A$$



#### MUO-RCOSIM MULTI-RATE GRAPH TRANSFORMATION (CONT'D)

#### Multi-Rate Graph Transformation Algorithm

- 1) Compute the hyper-step HS=lcm  $(H(o_1), H(o_2), ..., H(o_n))$
- 2) For each operation  $o_i$  in the graph
  - Compute the repetition factor r<sub>i</sub>= HS/ H(oi)
- 3) Repeat each operation o<sub>i</sub>, r<sub>i</sub> times
- 4) Add edges between successive occurrences of each operation
- 5) For each edge (o<sub>i</sub>,o<sub>j</sub>)
  - If  $H(o_i) > H(o_j)$  (slow to fast dependency)

Add edges  $(o_i^s, o_j^u)$ ,  $s \in \{1, 2, ..., r_i\}$ ,  $u = \left[s \times \frac{H(o_i)}{H(o_j)}\right]$ 

- If  $H(o_i) < H(o_j)$  (fast to slow dependency) Add edges  $(o_i^s, o_j^u), u \in \{1, 2, ..., r_j\}, s = \left| u \times \frac{H(o_j)}{H(o_j)} \right|$
- If  $H(o_i) = H(o_j)$ Add edges  $(o_i^s, o_j^u)$  between corresponding occurrences
- 6) For each FMU
  - Add edges between the occurrence s of the state operation and all the input and output operations of the next occurrence s+1
- 7) Stop when all operations and edges have been visited





#### MUO-RCOSIM MULTI-CORE SCHEDULING

- Off-line heuristic approach: Similar to SynDEx (INRIA) [Grandpierre et al., 1999]
- N operations, each one:
  - Computation time
  - Earliest and latest start and end dates  $\rightarrow$  Takes into account the synchronization cost
- Objective: Minimize the makespan (multiprocessor critical path) of the graph
- Cost function: Schedule pressure is the difference between:
  - Flexibility: Freedom degree of an operation: time interval inside which o<sub>i</sub> may be executed without increasing the makespan
  - Penalty: Critical path increase by setting an operation on a processor accounting for synchronization cost



#### MUO-RCOSIM MULTI-CORE SCHEDULING (CONT'D)

#### **Multi-core scheduling heuristic**

- 1) For each operation o<sub>i</sub>
  - Compute S<sub>i</sub> (resp. E<sub>i</sub>) the earliest start (resp. end ) time, and S'<sub>i</sub> (resp. E'<sub>i</sub>) the latest start (resp. end ) time
  - Compute the flexibility  $F_i = CP E_i E'_i$
- 2) Set  $\Omega$  the set of operations without predecessors
- 3) Repeat
  - For each pair (operation o<sub>i</sub> in Ω, core p<sub>j</sub>)
     Compute the increase (cost) of scheduling o<sub>i</sub> on p<sub>j</sub>
     Select for o<sub>i</sub>, the core p<sub>i</sub> which minimizes the cost of scheduling o<sub>i</sub>
  - Find the operation o<sub>i</sub> with the maximal cost on its selected core
  - Allocate o<sub>i</sub> to its selected core
  - Remove  $o_i$  from  $\Omega$
  - Add to  $\Omega$  every operation whose predecessors have been scheduled
  - Stop when all the operations have been scheduled



# TESTS

# Case study

- Spark Ignition RENAULT F4RT engine
- 6 FMUs, more than 100 operations
- Around 300 operations after applying the multi-rate transformation algorithm
  - Communication steps
    - Airpath/control: 100 μs
    - Cylinders: 20 µs
  - Integration step = communication step for all FMUs
- 3 approaches are compared
  - RCosim: Mono-rate, restricted allocations of the operation
  - MU-RCosim: MUlti-rate, restricted allocation of the operations
  - MUO-RCosim: MUlti-rate, uses the acyclic Orientation heuristic to handle mutual exclusion constraints





TESTS SPEED-UP

# • Speed-up = $\frac{Sequential execution time}{Parallel execution time}$

 Best speed-up close to 2.9 reached with 5 cores (compared to mono-core schedule)

- MUO-RCosim > MU-RCosim > RCosim
- Thanks to the mutual exclusion heuristic, an efficient execution order for mutual exclusive operations is defined
- This order tends to allow the multi-core scheduling heuristic to better adapt the potential parallelism to the execution platform



#### RCOSIM APPROACH ACCURACY: ELIMINATION OF DELAYS

• Torque is a direct feedthrough output: e.g. Y<sub>A3</sub>

- Expected delays with Standard Co-simulation (Std-Cosim) due to arbitrary order execution decision between models
- No delays with RCosim
  - The execution order is compliant with initial model





# OUTLINE

# Background

- Co-simulation: context & challenges
- Real-time Co-Simulation: an open problem
- Improving parallelism with the RCosim approach: Refined Co-simulation
- Ensuring co-simulation accuracy with CHOPtrey extrapolation approach
- Mapping real-time constraints for HiL
- Future work



# CONTEXT-BASED EXTRAPOLATION IMPROVE AGAIN THE SIMULATION ACCURACY

- Limitation: with RCosim, errors are reduced but still exist
- Reason: Input data is held constant during the communication step
- <u>Dilemma</u>: **/** ∕ communication step
  - ✓ Integration error
  - ↗↗ Speed-up
- Idea: Extrapolate input signals to
  - Enlarge intervals
  - Reduce simulation errors



Digicosme GT OVSTR - 26/04/2017

# **RELATED WORK**

# Difficulties

Related work on extrapolations treated the continuous case

- Successful for non-stiff systems / Encountered problems with stiff systems
- Complex systems with hybrid behavior is even more difficult to predict
  - Nonlinearities, discontinuities,...
- → Hard to predict the future behavior (from past observations)
  - Polynomial prediction fails due to the discontinuities
  - No universal prediction scheme, efficient with every signal
- Challenges: fast, causal and reliable prediction
  - Predictor computing cost << extra model computations with small communication steps</p>
  - Accurate predictions for any signal (blocky/smooth; slow/steep onsets)
- Idea: Borrow the context-based approach from lossless image encoders



# CHOPRED: COMPUTATIONALLY HASTY ONLINE PREDICTION CHOPOLY: CAUSAL HOPPING OBLIVIOUS POLYNOMIALS

# • $P_{\delta,\lambda,\omega}$ : least squares polynomial predictor

- $\delta$ : prediction degree;
- $\lambda$ : prediction frame length
- $\bullet \omega$ : weighting factor
- u: input signal;  $\tau$ : relative time for prediction
- Weighted moment:  $\overline{m}_{d,\lambda,\omega} = \sum_{l=0}^{\lambda-1} (\lambda l)^{\omega} l^d u_{-l}$
- Weighted sum of integer powers:  $\overline{z}_{d,\lambda,\omega} = \sum_{l=0}^{\lambda-1} (\lambda l)^{\omega} l^d$
- General formula for extrapolation:
  - Use of LUT → fast computation

$$u(\tau) = \begin{bmatrix} 1 & \tau & \cdots & \tau^{\delta} \end{bmatrix} \begin{bmatrix} z_{0,\lambda,\omega} & z_{1,\lambda,\omega} & \cdots & (-1) & z_{0,\lambda,\omega} \\ -\overline{z}_{1,\lambda,\omega} & & \vdots \\ \vdots & & \vdots \\ (-1)^{\delta} \overline{z}_{\delta,\lambda,\omega} & \cdots & \cdots & \overline{z}_{2\delta,\lambda,\omega} \end{bmatrix} \begin{bmatrix} \overline{m}_{0,\lambda,\omega} \\ -\overline{m}_{1,\lambda,\omega} \\ \vdots \\ (-1)^{\delta} \overline{m}_{\delta,\lambda,\omega} \end{bmatrix}$$

 $\begin{bmatrix} \overline{z}_{0}, \alpha & -\overline{z}_{1}, \alpha & \cdots & (-1)^{\delta} \overline{z}_{\delta}, \alpha \end{bmatrix}^{-1}$ 

#### CHOPATT: CONTEXTUAL AND HIERARCHICAL ONTOLOGY OF PATTERNS META- OR DECISIONAL CONTEXT SELECTION

- Worst case scenario without extrapolation:  $\Delta_{\text{worst}} = |u_0 u_{-1}|$
- Best prediction pattern:  $\Delta_{\text{best}} = \min_{\omega \in \Omega} |u_0 \hat{u}_{-1}^{\omega}|$ ;  $\Omega = \{0, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2\}$ • Ratio:  $\rho = \frac{\Delta_{\text{best}}}{\Delta_{\text{worst}}}$ • Threshold:  $0.7 \leq \Gamma < 1$  e.g.  $\Gamma = 90\%$
- If  $\rho > \Gamma$  then sharp and fast variation  $\rightarrow$  Select the decisional context: cliff context





#### CHOPATT: CONTEXTUAL AND HIERARCHICAL ONTOLOGY OF PATTERNS FUNCTIONAL CONTEXT SELECTION





#### CHOPATT: CONTEXTUAL AND HIERARCHICAL ONTOLOGY OF PATTERNS FUNCTIONAL CONTEXT SELECTION

n(ame)	#	<i>d</i> <sub>-1</sub>	<i>d</i> <sub>0</sub>	<i>d</i> <sub>-1</sub> . <i>d</i> <sub>0</sub>	(δ, λ, ω)
f(lat)	0	0	0	0	(0, 1, .)
c(alm)	1	<i>C</i> <sub>1</sub>	<i>C</i> <sub>2</sub>	any	(2, 5, .)
m(ove)	2	<i>C</i> <sub>1</sub>	$\bar{C}_2$	any	(0, 1, .)
r(est)	3	$\bar{C}_1$	C <sub>2</sub>	any	(0, 2, .)
t(ake)	4	$\bar{C}_1$	$\bar{C}_2$	>0	(1, 3, .)
j(ump)	5	$\bar{C}_1$	$\bar{C}_2$	< 0	(0, 1, .)





**IFPEN Transports Energie** 

# SIMULATION RESULTS WITH CHOPtrey AUTOMATIC DETECTION OF SHARP VARIATION

#### Same case study

- 118 states/312 events
- Solver: LSODAR
- Communication step: 200µs
- Conventional 1<sup>st</sup> & 2<sup>nd</sup> order extrapolation
  - Fails on the engine model
  - Major causes:
    - Discontinuities
    - Sharp variations



#### SIMULATION RESULTS WITH CHOPtrey AUTOMATIC DETECTION OF SHARP VARIATION

# Same case study

- 118 states/312 events
- Solver: LSODAR
- Communication step: 200µs
- Conventional 1<sup>st</sup> & 2<sup>nd</sup> order extrapolation
  - Fails on the engine model
  - Major causes:
    - Discontinuities
    - Sharp variations
- ➔ Context-based extrapolation?







30

#### SIMULATION RESULTS WITH CHOPtrey AUTOMATIC DETECTION OF SHARP VARIATION

#### Same case study

- 118 states/312 events
- Solver: LSODAR
- Communication step: 200µs
- Conventional 1<sup>st</sup> & 2<sup>nd</sup> order extrapolation
  - Fails on the engine model
  - Major causes:
    - Discontinuities
    - Sharp variations
- ➔ Context-based extrapolation?





IFPEN Transports Energi

# SIMULATION RESULTS WITH CHOPtrey AUTOMATIC SELECTION OF THE WEIGHTING FACTOR

- No unique best weighting factor ω due to complex coupled systems
- $\rightarrow$  Dynamic selection of  $\omega$ 
  - At each communication step,  $\omega_{\text{best}}$  is selected and used for the current step
  - $\rightarrow$ Cumulative integration error is the lowest one





#### CHOPtrey PERFORMANCE SPEED-UP VERSUS ACCURACY

• The speed-up factor is compared with single-threaded reference

- The model is split into 5 threads integrated in parallel on 5 cores
  - Containment of events detection handling → solvers accelerations → overcompensate multi-threading costs
- The relative error variation is compared with ZOH at 100 µs

Communication step	Prediction	Speed-up factor	Relative error variation (%)		
			Burned gas density	Fuel density	
100 µs	ZOH	8.9 +12.5%	-	-	
<b>2</b> 50 μs	ZOH	10.01	7	341	
	CHOPtrey	10.07	-26	21	



# OUTLINE

# Background

- Co-simulation: context & challenges
- Real-time Co-Simulation: an open problem
- Improving parallelism with the RCosim approach: Refined Co-simulation
- Ensuring co-simulation accuracy with CHOPtrey extrapolation approach
- Mapping real-time constraints for HiL
- Future work



FPEN Transports Energ

## **REAL TIME SIMULATION** FROM REAL TIME TO SIMULATED TIME



Absolute Simu constraints with offset  

$$R_v, D_v = (\Phi, T)$$
offset
period

$$n \times T_{IO} = t_k$$

$$O \longrightarrow u \quad x \quad y \longrightarrow I$$

$$DF \Leftrightarrow u_k \rightarrow y_k$$

- Propagation *r-simu*  $R_u = R_O$ 
  - $R_x$ : offset h
  - $DF \Longrightarrow R_y = R_u$
  - •



Absolute Simu constraints with offset  

$$R_v, D_v = (\Phi, T)$$
offset
period

$$n \times T_{IO} = t_k$$

$$O \longrightarrow u \quad x \quad y \longrightarrow I$$

$$NDF \Leftrightarrow u_k \not\rightarrow y_k$$

- Propagation *r-simu*  $R_u = R_O$ 
  - $R_x$ : offset h
  - $DF \Longrightarrow R_y = R_u$
  - $NDF \Longrightarrow R_y$  : offset h



Absolute Simu constraints with offset  

$$R_v, D_v = (\Phi, T)$$
offset
period

$$n \times T_{IO} = t_k$$

$$O \longrightarrow u \quad x \quad y \longrightarrow I$$

$$DF \Leftrightarrow u_k \rightarrow y_k$$

- Propagation *r-simu*  $R_u = R_O$ 
  - $R_x$ : offset h
  - $DF \Longrightarrow R_y = R_u$
  - $NDF \Longrightarrow R_y$  : offset h

- Propagation *d-simu*  $D_y = D_I$ 
  - $D_x$ : no offset
  - $DF \Longrightarrow D_u = D_y$
  - ٠



Absolute Simu constraints with offset 
$$R_v, D_v = (\Phi, T)$$
  
offset period

$$n \times T_{IO} = t_k$$

$$O \longrightarrow u \quad x \quad y \longrightarrow I$$

$$NDF \Leftrightarrow u_k \not\rightarrow y_k$$

- Propagation *r-simu*  $R_u = R_O$ 
  - $R_x$ : offset h
  - $DF \Longrightarrow R_y = R_u$
  - $NDF \Longrightarrow R_y$  : offset h

- Propagation *d-simu*  $D_y = D_I$ 
  - $D_x$ : no offset
  - $DF \Longrightarrow D_u = D_y$
  - $NDF \Rightarrow D_u$ : offset -h



# EXAMPLE OF PROPAGATION RELATIVE MESHES



$(\Phi,T)$ $v$	A.u	A.x	A.y = B.u	B.x	<b>B</b> .y
$R_{_{\mathcal{V}}}$	$R_o = (0,2)$	(1,2)	(1,2)	(2,2)	(1,2)
$D_v$	(-1,2)	(0,2)	(0,2)	(0,2)	$D_I = (0,2)$





$$\forall D_{v} = (\Phi, T), \ d_{v,i} = \left[\frac{h \times i - \Phi}{T}\right] \times T$$

$$\forall R_{v} = (\Phi, T), \ r_{v,i} = \left\lfloor\frac{h \times i - \Phi}{T}\right\rfloor \times T$$

41

Digicosme GT OVSTR - 26/04/2017





43





# FUTURE WORK

#### Short term

- Comparison of RCOSIM heuristics with an exact scheduling algorithm
- Comparison of RCOSIM offline approach with on-line scheduling

#### Mid term

- Extension of real-time constraints propagation rules to the RCOSIM fine-grained approach
- Real-time multicore scheduling heuristics for an HiL version of the RCOSIM



#### REFERENCES

 S. Saidi et al., "Automatic Parallelization of Multi-Rate FMI-based Co-Simulation on Multi-Core" Spring Simulation Multiconference (Apr. 2017)

 A. Ben Khaled-El Feki et al., "CHOPtrey: contextual online polynomial extrapolation for enhanced multi-core co-simulation of complex system", Simulation, 2017, vol. 93(3), pp. 185-200. DOI: 10.1177/0037549716684026

 A. Ben Khaled-El Feki et al., "Fast multi-core co-simulation of Cyber-Physical Systems: application to internal combustion engines", *Simulation Modelling Practice and Theory*, 2014, vol. 47, pp. 79-91. <u>DOI : 10.1016/J.SIMPAT.2014.05.002</u>

