

# Design methods and tools for real-time (automotive) embedded systems

*Abhijit Davare, **Marco Di Natale**, Paolo Giusto, Claudio Pinello, Alberto Sangiovanni-Vincentelli, Wei Zheng, Qi Zhu,*

GM Research and Development  
University of California at Berkeley,  
Scuola Superiore S. Anna,

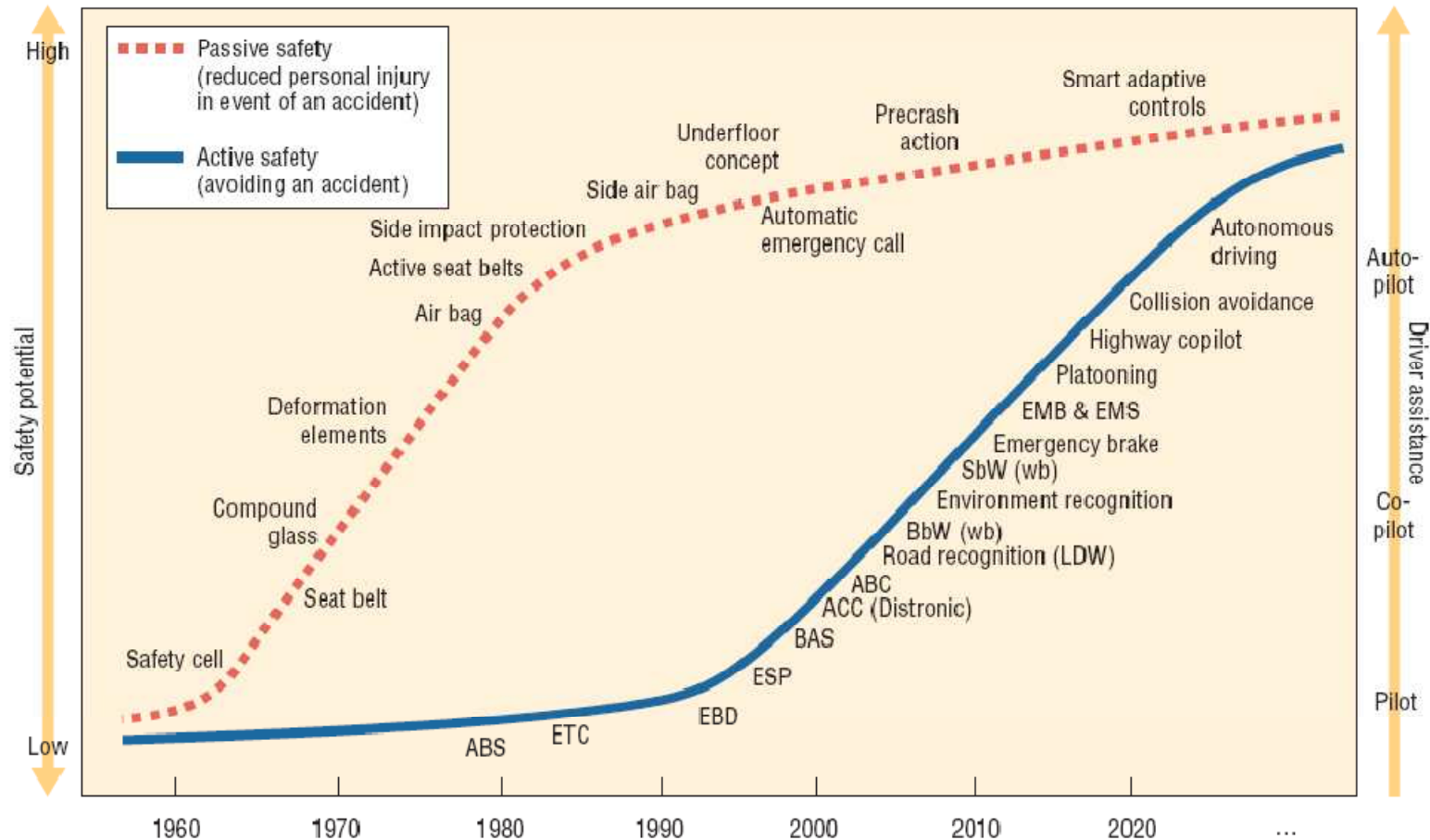




# Outline

- Automotive architecture trends and challenges
- Platform-based system-level design and timing evaluation metrics
- Issues with model-based design
- From analysis to synthesis
- Activation models and end-to-end latencies
- Problem definition
  - Example
- MILP Optimization
- Case Study

# Active and Passive Safety



ABC	Active body control	EBD	Electronic brakeforce distribution
ABS	Antilock brake system	EMB	Electromechanical brakes
ACC	Adaptive cruise control	EMS	Electromechanical steering
BAS	Brake assist system	ESP	Electronic stability program
BbW	Brake by wire	ETC	Electronic traction control
CA	Collision avoidance	SbW	Steer by wire
DbW	Drive by wire	(wb)	with mechanical backup

## AS - ACC (from Continental web site)

- Adaptive Cruise Control (ACC) – Chassis Electronics Combined with Safety Aspects

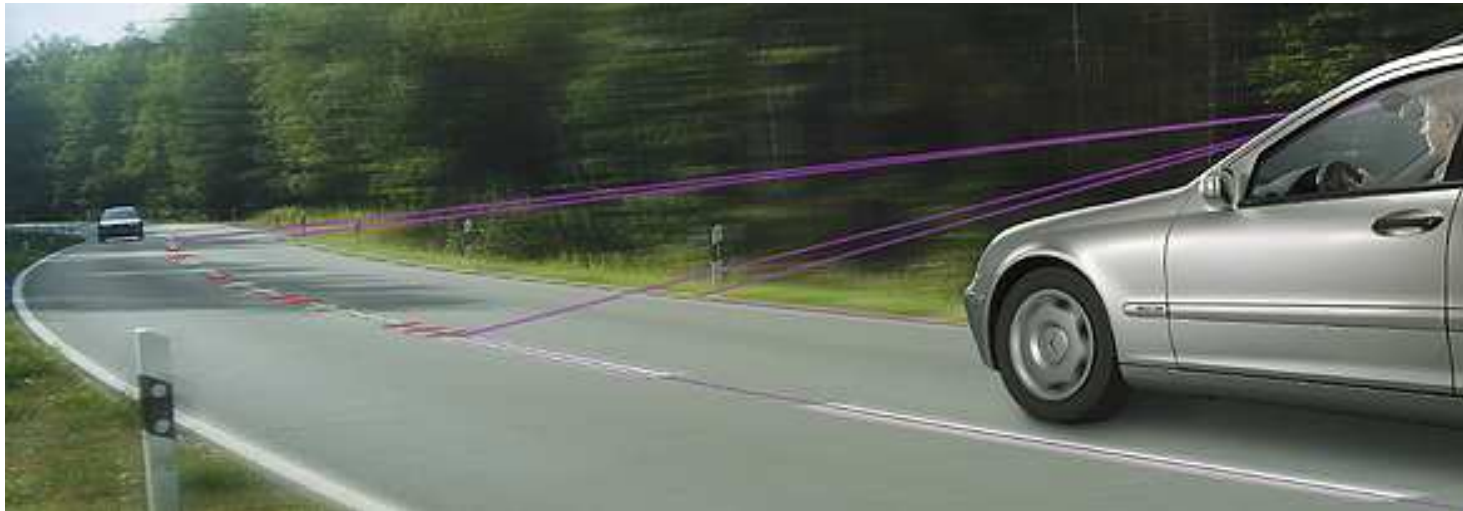


As with conventional cruise control, the driver specifies the desired velocity - ACC consistently maintains this desired speed.

In addition, the driver can enter the desired distance to a vehicle driving in front. If the vehicle now approaches a car travelling more slowly in the same lane, ACC will recognize the diminishing distance and reduce the speed through intervention in the motor management and by braking with a maximum of 0.2 to 0.3 g until the preselected distance is reached. If the lane is clear again, ACC will accelerate to the previously selected desired tempo.

## AS-LDW (from Continental web site)

- Lane Departure Warning System (LDW)



LDW will warn the driver if he or she is on the verge of inadvertently drifting out of the lane. Using a CMOS Camera and an image processing algorithm, this driver assistance system registers the course of the lane in relation to the vehicle. The system "sees", as it were, the course of the road and where the car is going. If the warning algorithm detects an imminent leaving of the current driving lane, the system warns the driver with haptic, kinesthetic, or acoustical feedback. Possible warning alerts can be a trembling in the steering wheel, a vibrating seat or a virtual washboard sound. Series production is planned for 2005.

# Evolution of Integrated Functions

Subsystem	Brake	HVAC	Body	Steering	Suspension	Object detection	Environm. sensing	Infotainm.	Occ. protection	Exterior lighting	Occupant Informatio	Engine	Transmissio	Telematics
<b>Post-2014</b>	function17													
	function16													
	function15													
	function14													
<b>to 2012/14</b>	function13													
	function12													
	function11													
	function10													
<b>to 2010/12</b>	function9													
	function8													
	function7													
	function6													
	function5													
<b>Pre-2004</b>	ACC													
	Stabilitrak 2													
	Onstar emergency notification													
	Speed-dependant volume													





## Automotive architecture trends

- Horizontally-integrated functions are becoming key differentiators and are gaining increasing authority
- An increasing number of functions will be distributed on a decreasing number of ECUs and enabled through an increasing number of smart sensors and actuators
  - today: > 5 buses and > 30 ECUs
- 90% of innovation in cars for the foreseeable future will be enabled through the Electronic Vehicle Architecture
- Transition from single-ECU Black-box based development processes to a system-level engineering process
  - System-level methodologies for quantitative exploration and selection,
  - From Hardware Emulation to Model Based Verification of the System
- Architectures need to be defined years ahead of production time, with incomplete information about (future) features
- Multiple non-functional requirements can be defined <sup>7</sup>

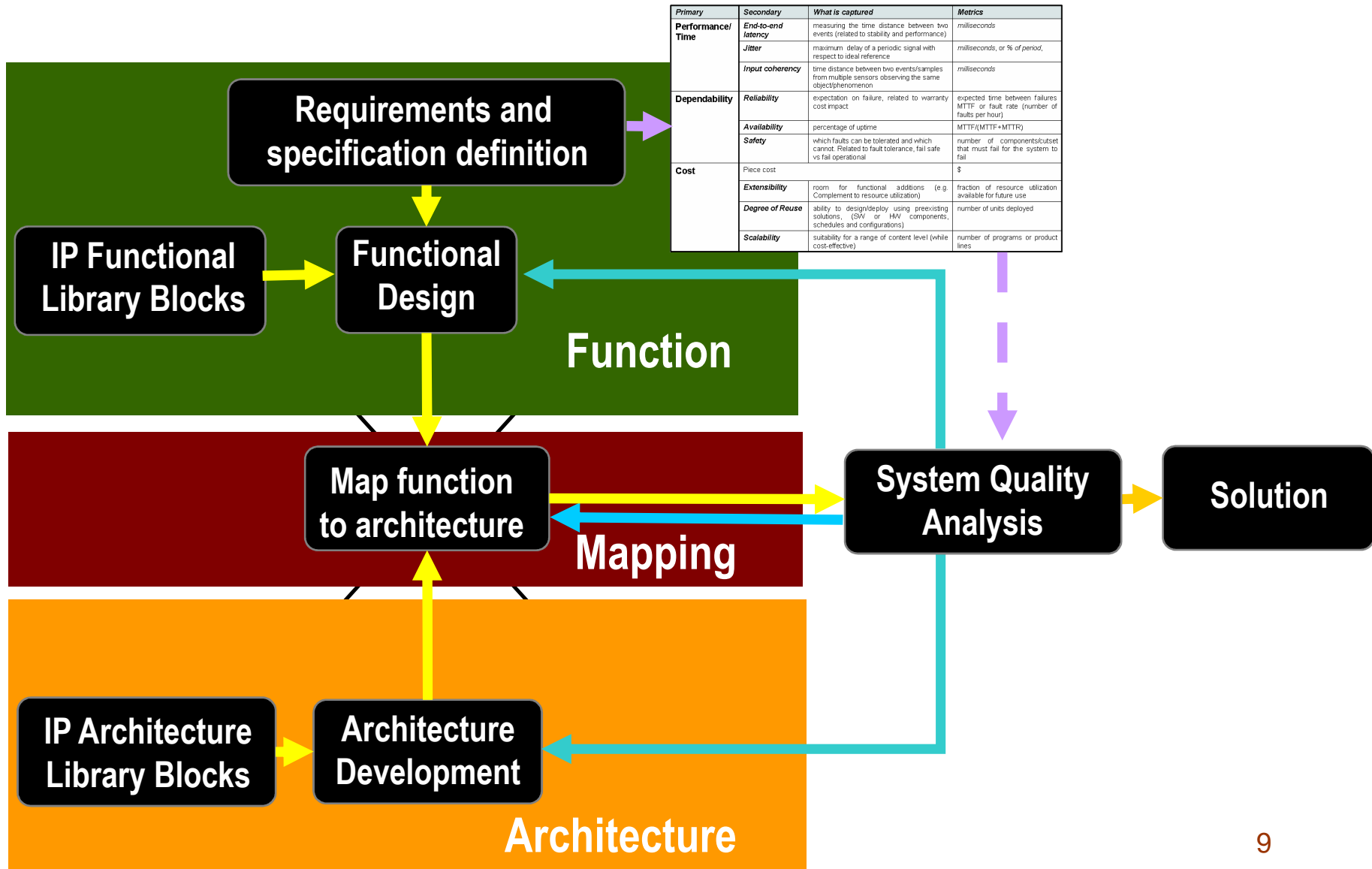


# Outline

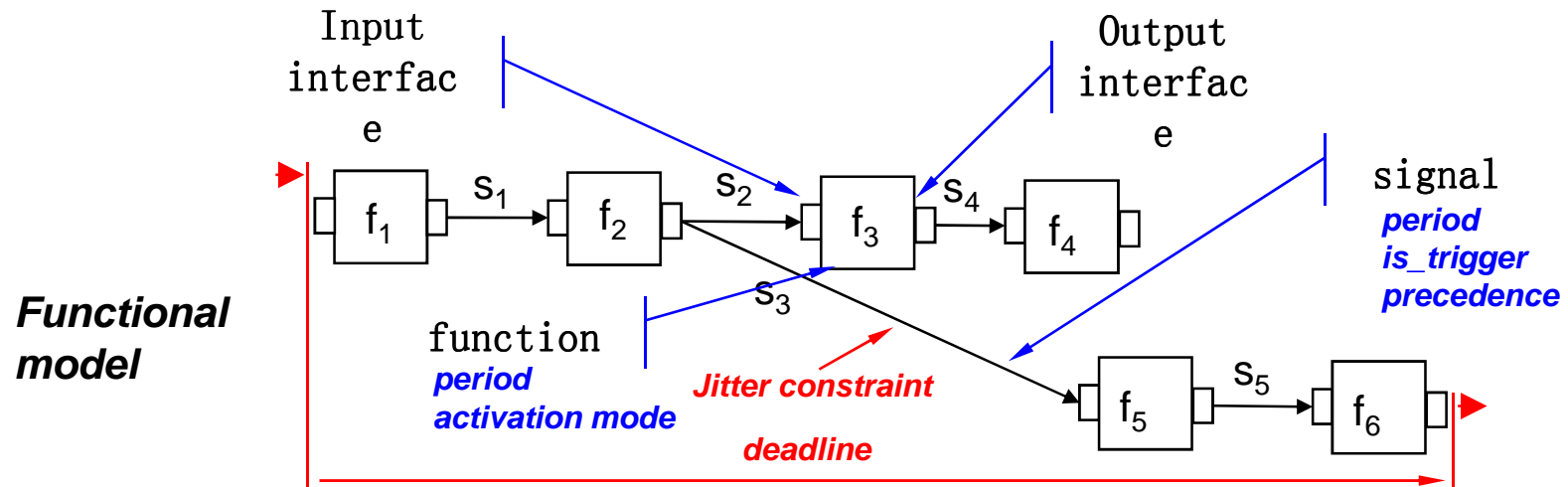
- Automotive architecture trends and challenges
- Platform-based system-level design and timing evaluation metrics
  - worst-case analysis
  - stochastic analysis
- Issues with model-based design
- From analysis to synthesis
- Activation models and end-to-end latencies
- Problem definition
  - Example
- MILP Optimization
- Case Study



# Deployment Design Process

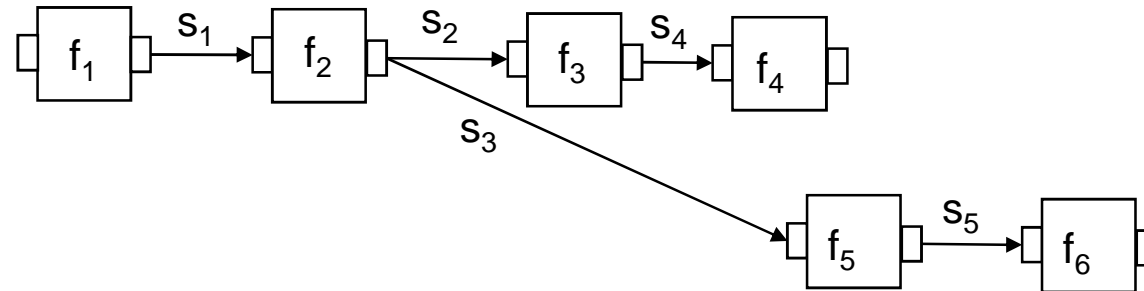


# Functional model

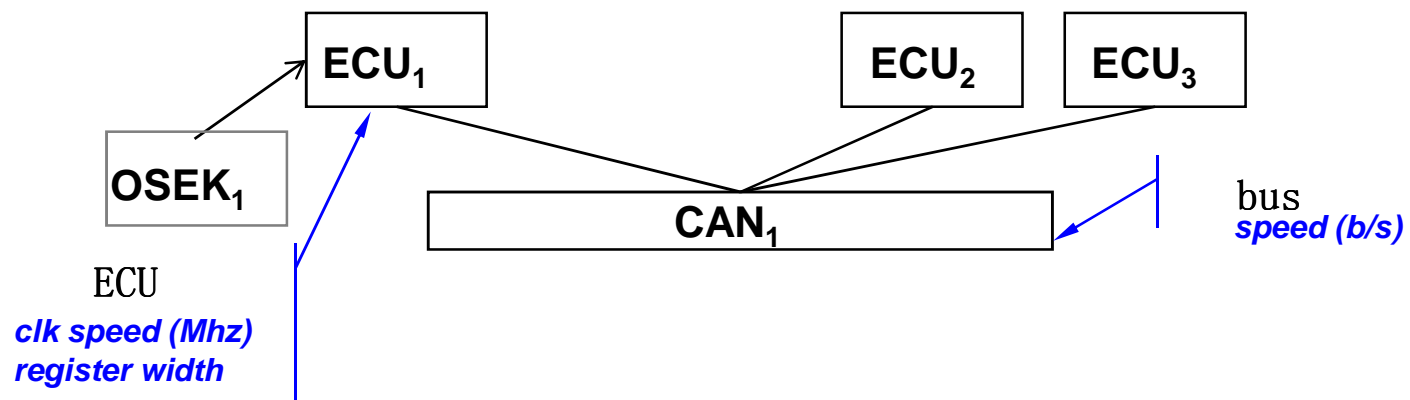


# Architecture model

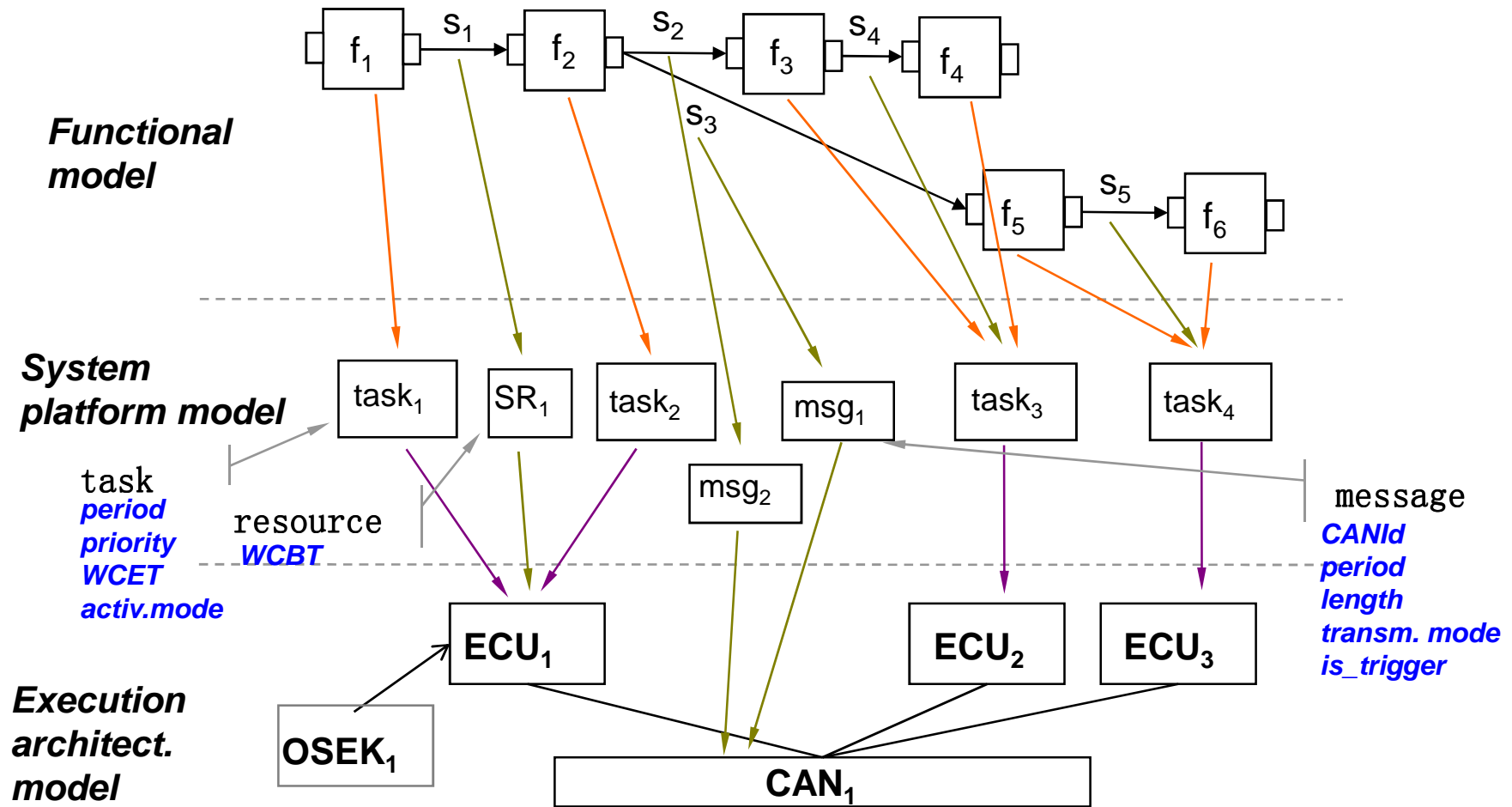
*Functional model*



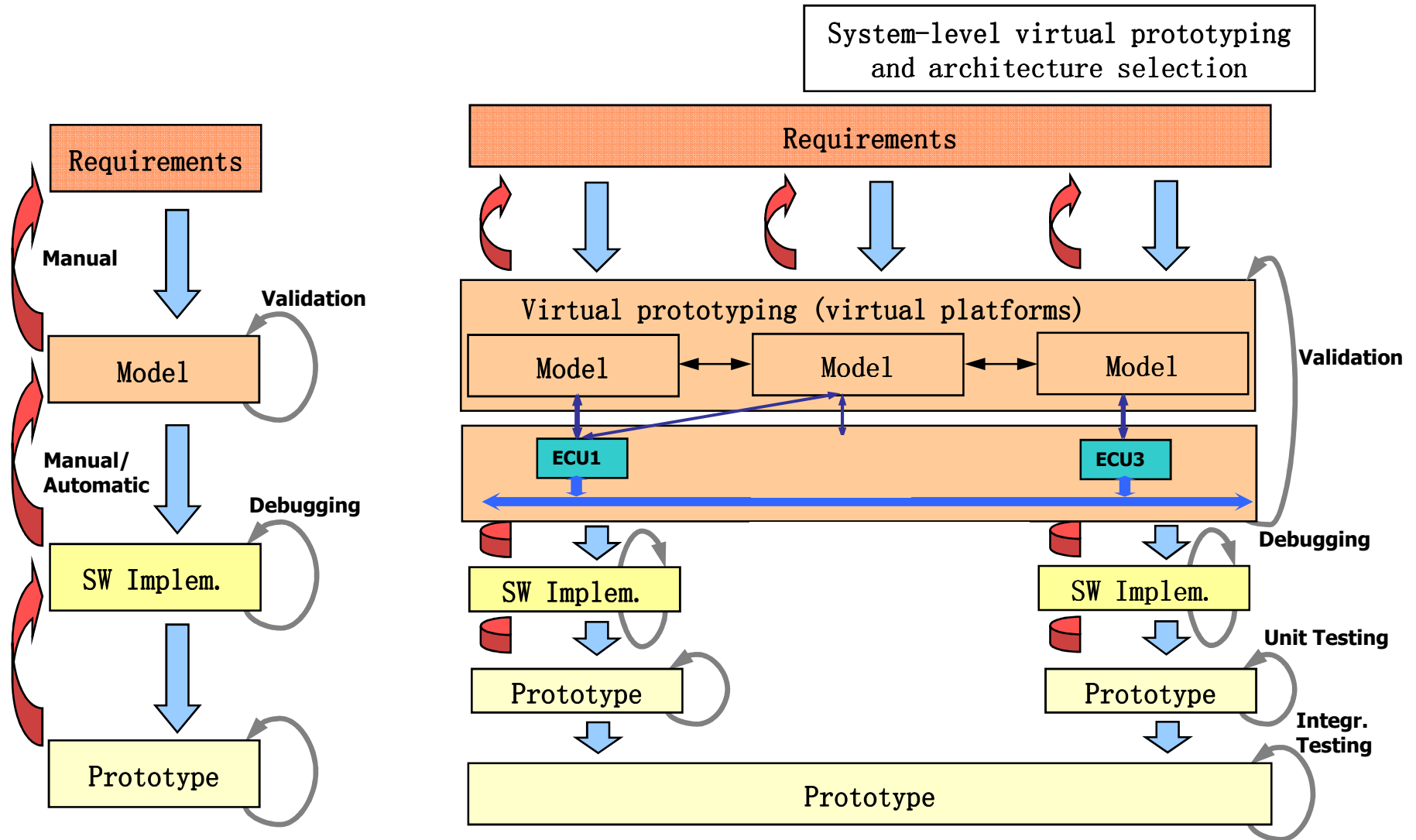
*Execution architect. model*



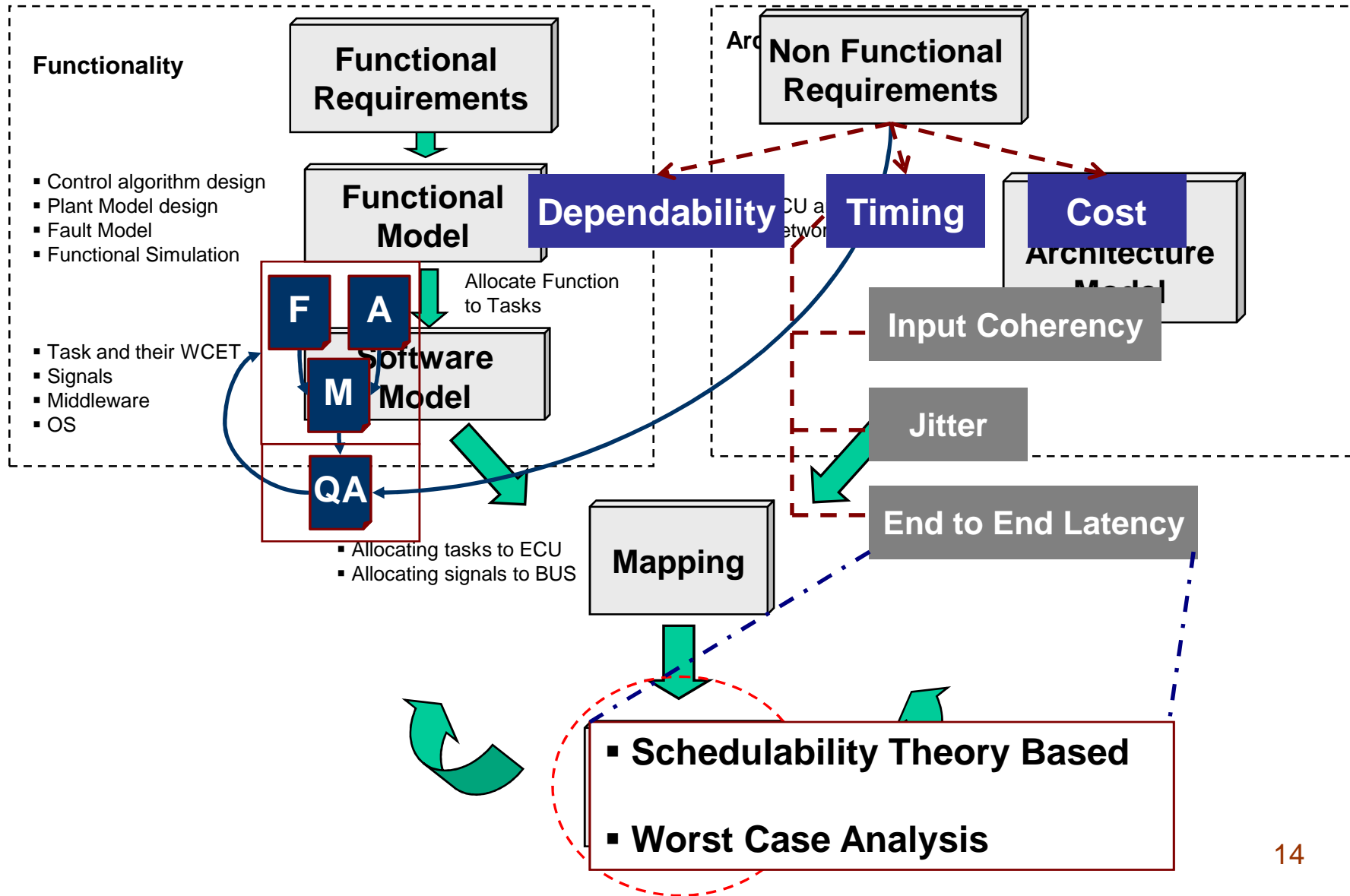
# Deployment model



# Tool integration platform



# Design Process and Requirement

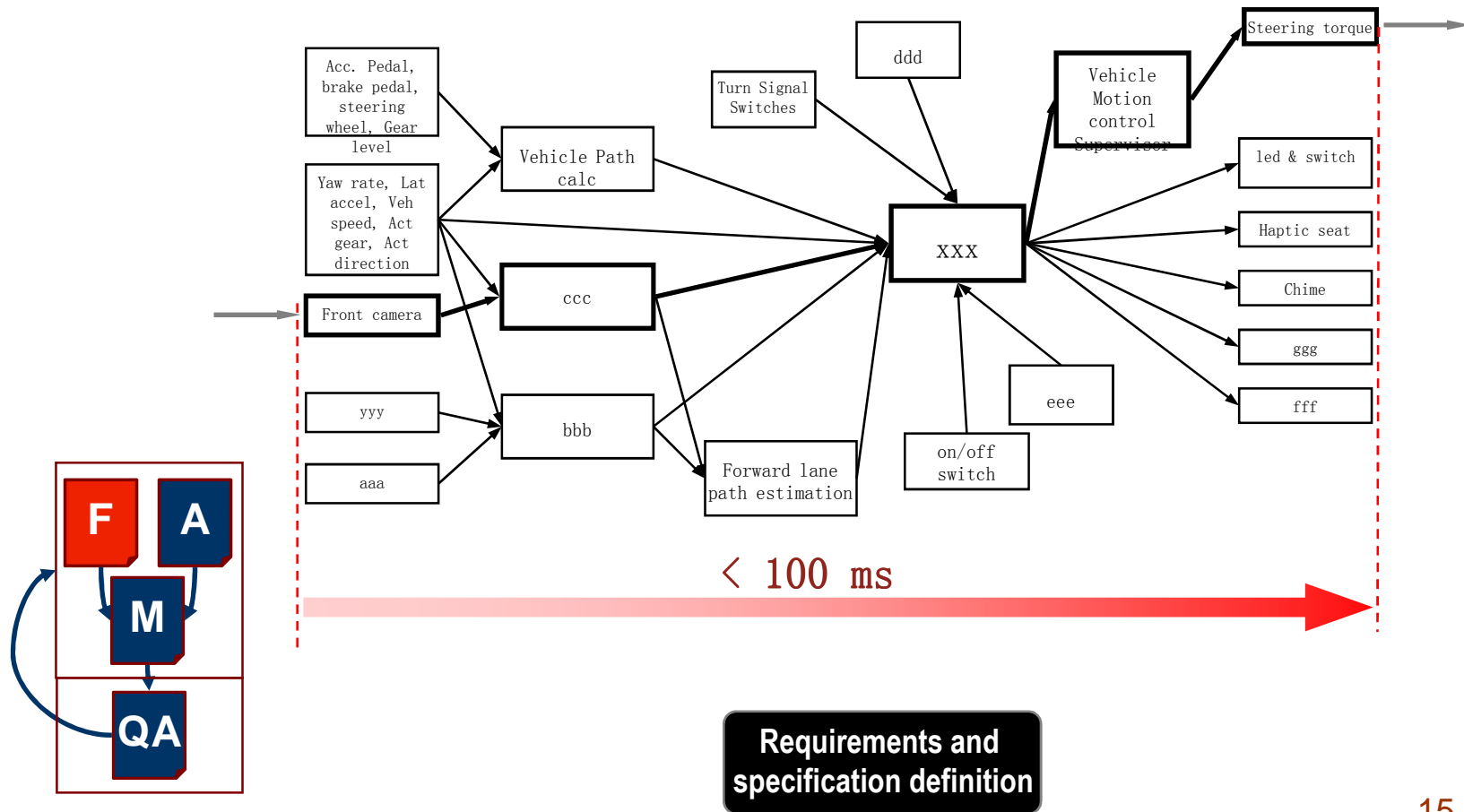


# Functional Model: An example

Function Example

XXX

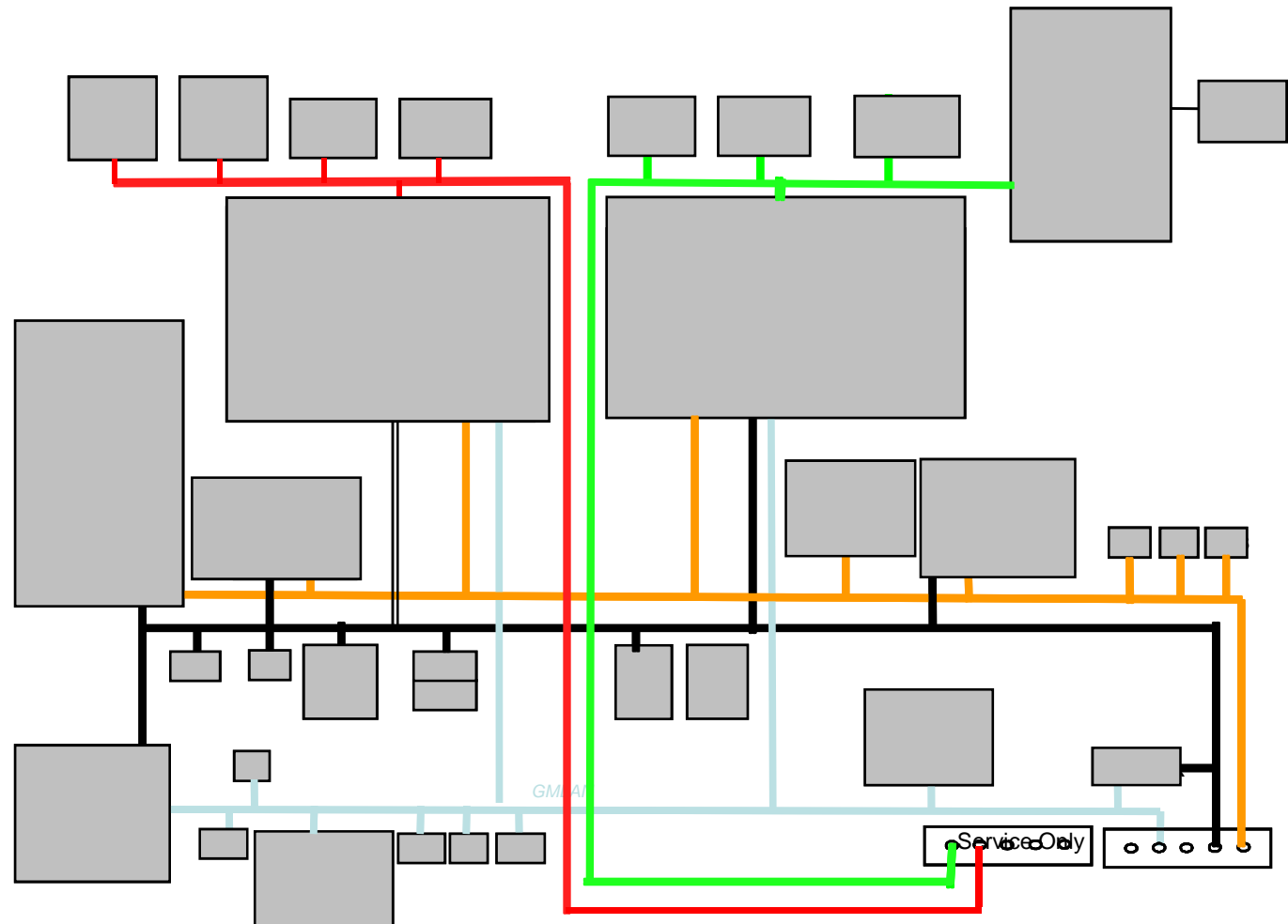
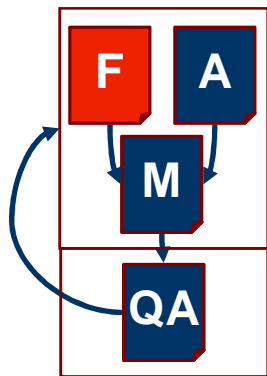
Functional Design



# Architecture Model: An example

Architecture  
Option

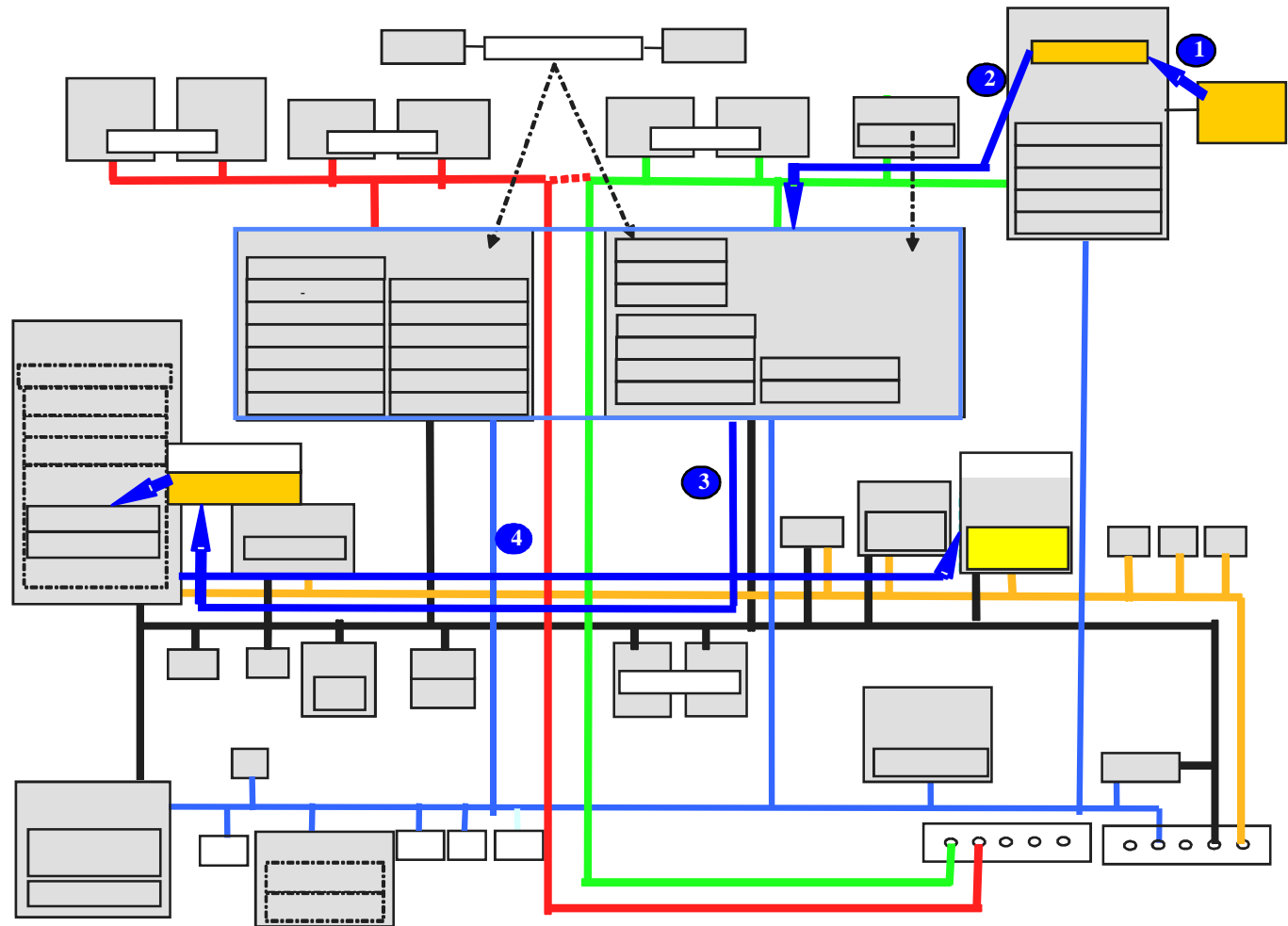
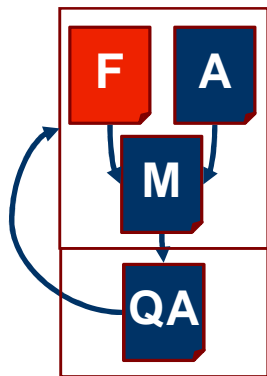
Architecture  
Development



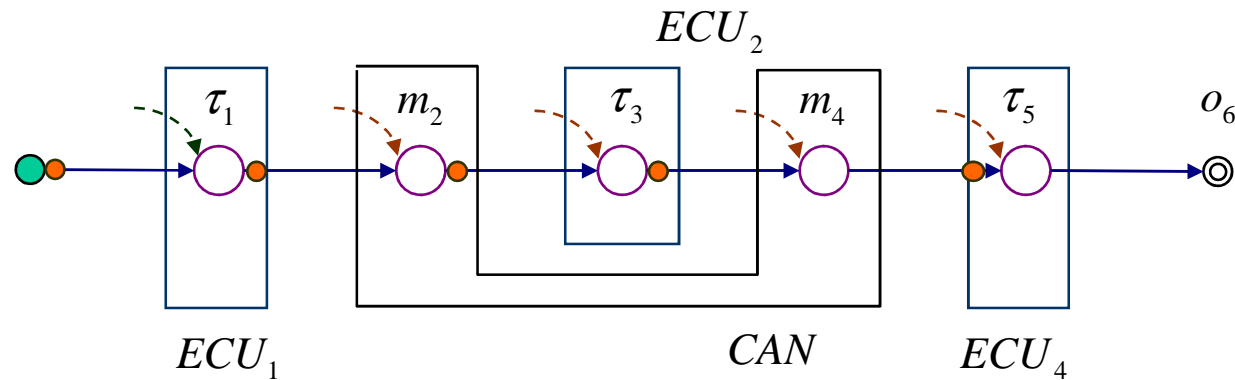


# Deployment: An example

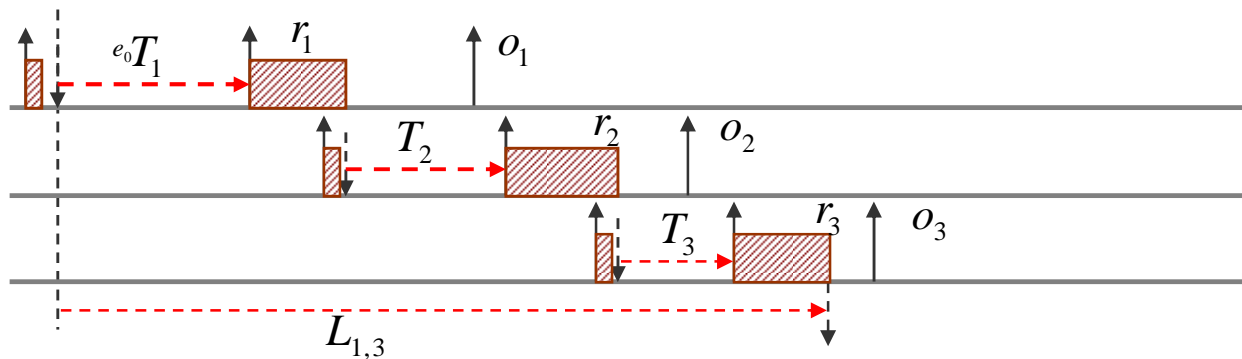
End-to-end latencies  
ECU and bus utilizations



# Periodic Activation Model



- Predictable activation model easy latency computation
- Suffers from high worst case latencies



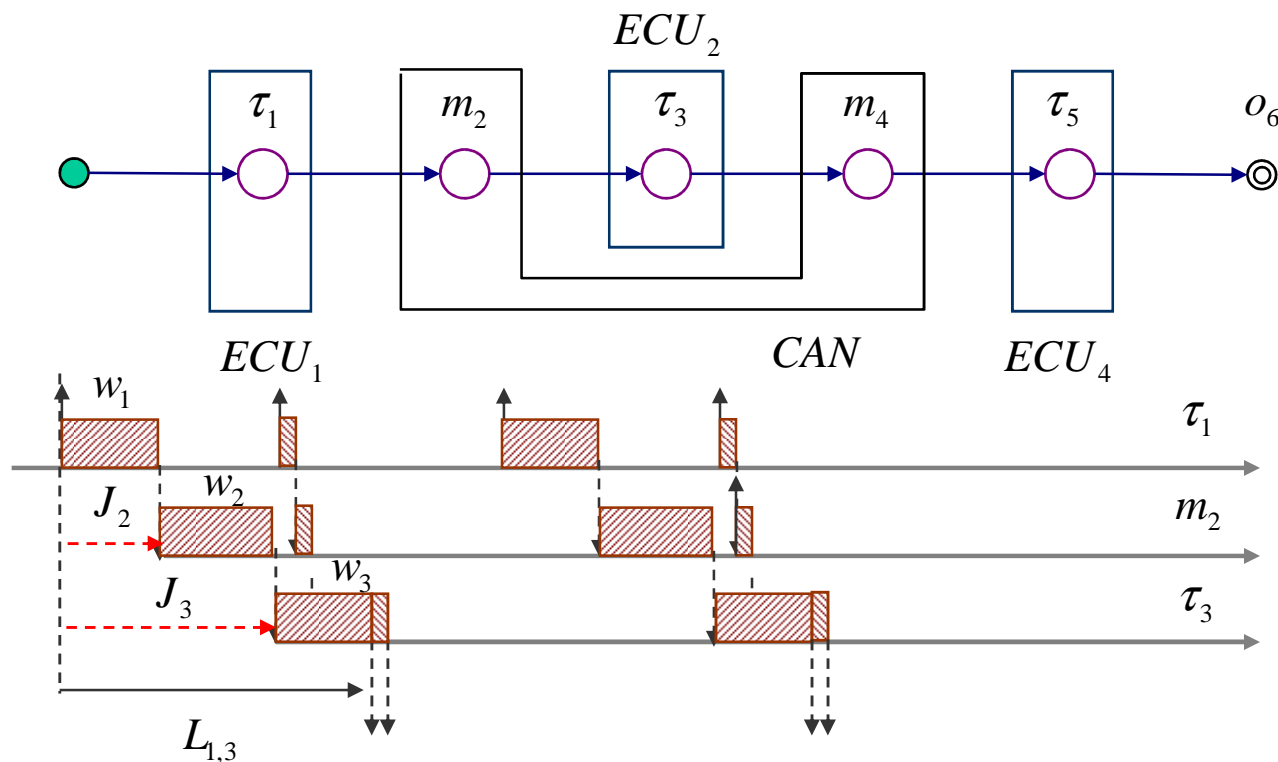
$$L_{i,j} = \sum_{k:o_k \in P(i,j)} (T_k + r_k)$$

Where

$$r_i = C_i + \sum_{j \in hp(i)} \left\lfloor \frac{r_i}{T_j} \right\rfloor C_j$$

$$L_{1,3} = T_1 + r_1 + T_2 + r_2 + T_3 + r_3$$

# Data Driven Activation Model



- Shorter end to end latencies
- Large interference intervals with bursty activations

$$L_{i,j} = \sum_{k:o_k \in P(i,j)} w_k$$

Where Approx.

$$w_i = C_i + \sum_{j \in hp(i)} \left\lceil \frac{w_i + J_j}{T_j} \right\rceil C_j$$

$$L_{1,3} = w_1 + w_2 + w_3$$

# Case study 1

Functions	Reqmt	Alt 1	Alt 2	Alt 4	Alt 4exp
function5	180	433.92 178.92	159.08 116.58	312.32 119.82	312.32 119.82
function4	100	195.21 155.21	109.35 89.35	180.93 70.93	180.93 70.93
function3	300	178.72 196.72	11.25 191.25	11.60 191.60	24.18 204.18
function2	300	520.99 139.99	479.06 129.06	479.19 129.19	489.19 139.19
function1	300	695.38 167.38	145.75 195.75	716.10 196.10	728.68 208.68

Synthesis  
opportunity

Functions	Reqmt	Alt 5	Alt 5exp	Alt 6	Alt 6 (event)
function5	180	310.58 118.08	310.58 118.08	230.06 72.56	130.1 60.06
function4	100	180.97 70.97	180.97 70.97	180.97 70.97	180.97 58.47
function3	300	162.74 162.74	162.74 162.74	162.74 162.74	163.9 123.9
function2	300	489.57 139.57	489.57 139.57	489.57 139.57	303.8 113.8
function1	300	537.24 167.24	537.24 167.24	537.24 167.24	318.9 128.9

- By transmitting messages "on event", the worst case latency can be reduced in most cases
- By properly allocating functions to ECUs the end-2-end latency can be improved



## Stochastic and simulation-based analysis

- Simulation
  - Built C++ simulator for can message analysis (at bit level – only arbitration)
  - Currently being expanded to end-to-end computations, periodic sampling model for latency analysis
- Stochastic analysis
  - Approximate analysis of pmf of message latencies in CAN bus (complete - target ?)
  - Future work
    - End-to-end analysis of sampling model
    - Regression-based analysis to define pmf from general information (such as load or loads at harmonic rates)

# Stochastic and simulation-based analysis

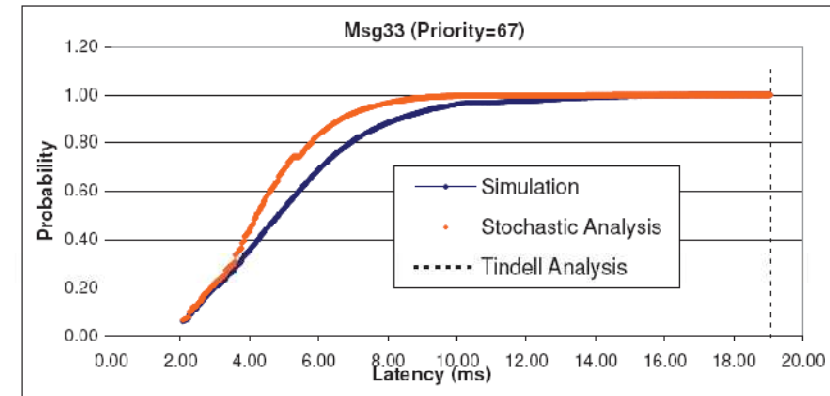
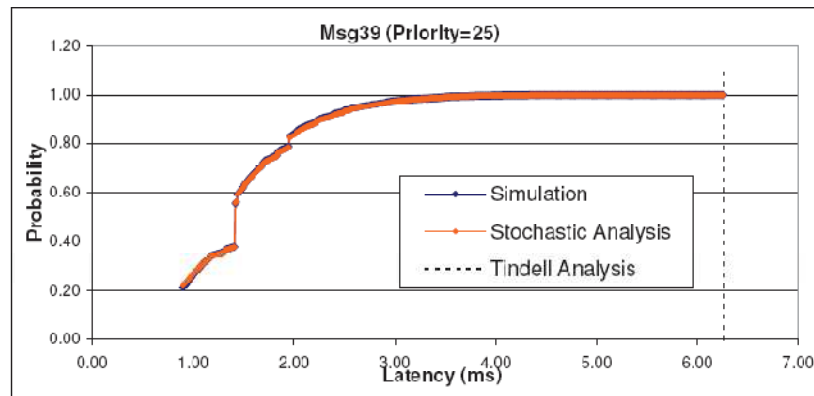
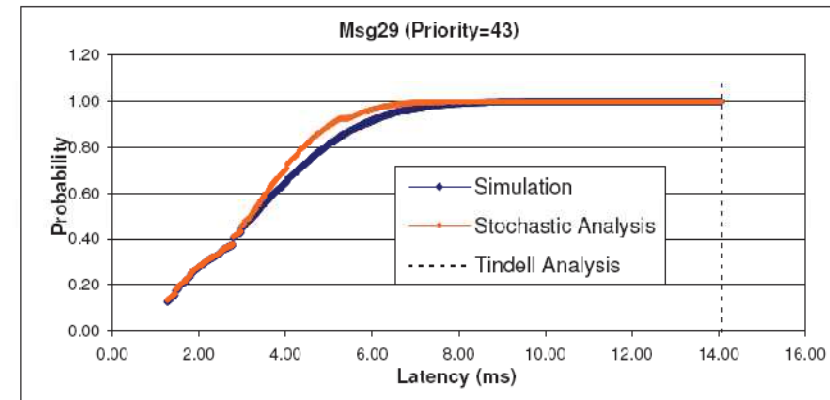
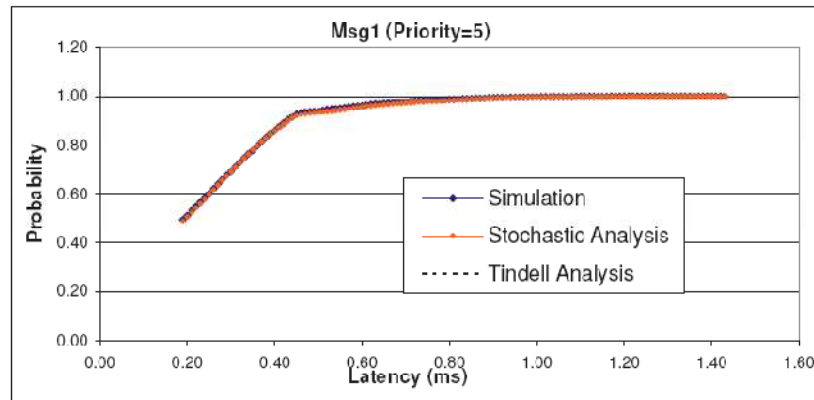


Figure 5. Latency *cdfs* of two high priority representative messages in the test set

Figure 6. Latency *cdfs* of two low priority representative messages in the test set

62 msg set (subset of chassis bus). Low priority msg – Distributions of latencies



# Outline

- Automotive architecture trends and challenges
- Platform-based system-level design and timing evaluation metrics
  - worst-case analysis
  - stochastic analysis
- **Issues with model-based design**
- From analysis to synthesis
- Activation models and end-to-end latencies
- Problem definition
  - Example
- MILP Optimization
- Case Study



## Issues with model-based development

- Model-based design methodologies
  - improve the quality and the reusability of software.
  - The possibility of defining components (subsystems) at higher levels of abstraction and with well defined interfaces allows separation of concerns and improves modularity and reusability.
  - The availability of verification tools (often by simulation) gives the possibility of a design-time verification of the system properties.
- However, most modern tools for model-based design have a number of shortcomings





## Issues with model-based development

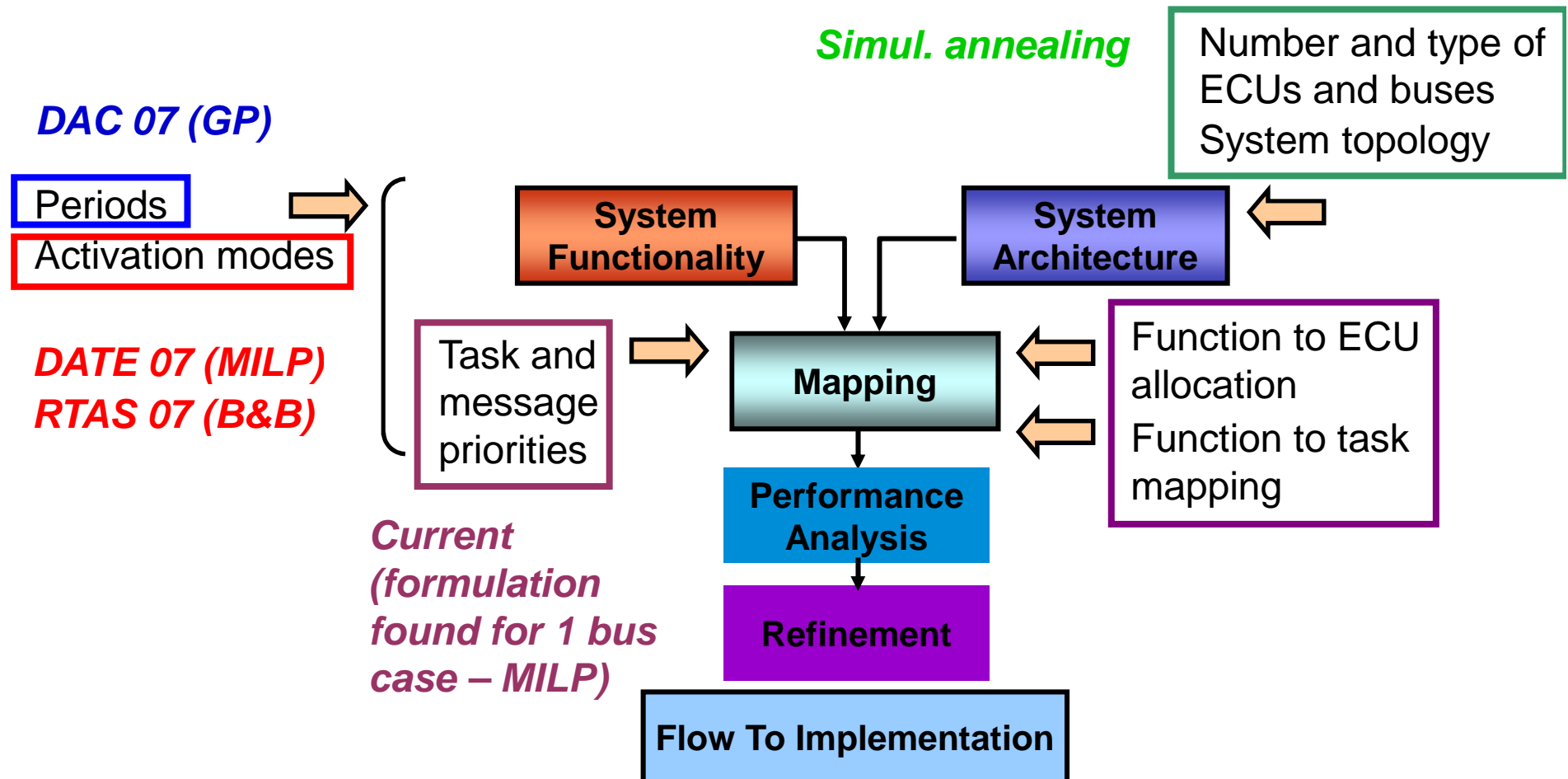
- *Lack of separation between the functional model and the architecture model*
- *Lack of support for the definition of the task and resource model*
- *Insufficient support for the specification of timing constraints and attributes*
- *Lack of modeling support for the analysis and the back-annotation of scheduling-related delays*
- *Issue of semantics preservation*



# Outline

- Automotive architecture trends and challenges
- Platform-based system-level design and timing evaluation metrics
  - worst-case analysis
  - stochastic analysis
- Issues with model-based design
- Time predictability and timing isolation
- **From analysis to synthesis**
- Activation models and end-to-end latencies
- Problem definition
  - Example
- MILP Optimization
- Case Study

# Opportunities for synthesis

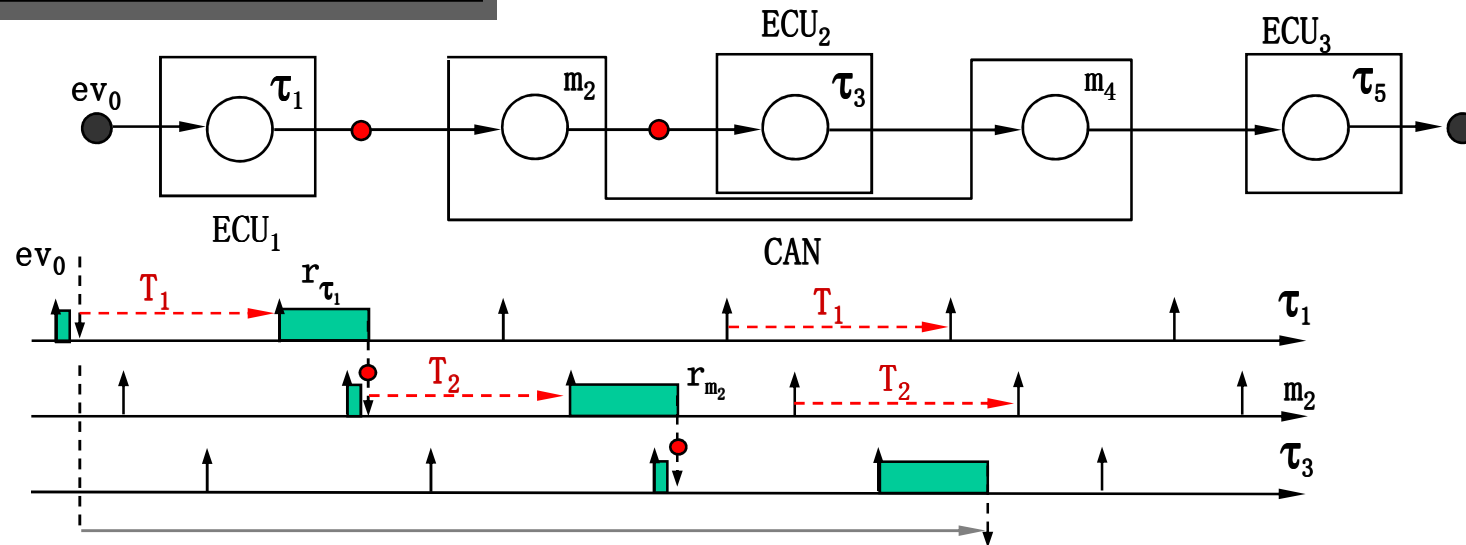


# Periodic Activation Model

High latency, but allows decoupling the scheduling problem

End-to-end latency analysis

Periodic asynchronous activation model



$$l_{(i,j)} = \sum_{k: o_k \in P(i,j)} (T_k + r_k) \quad \text{where (approx.)}$$

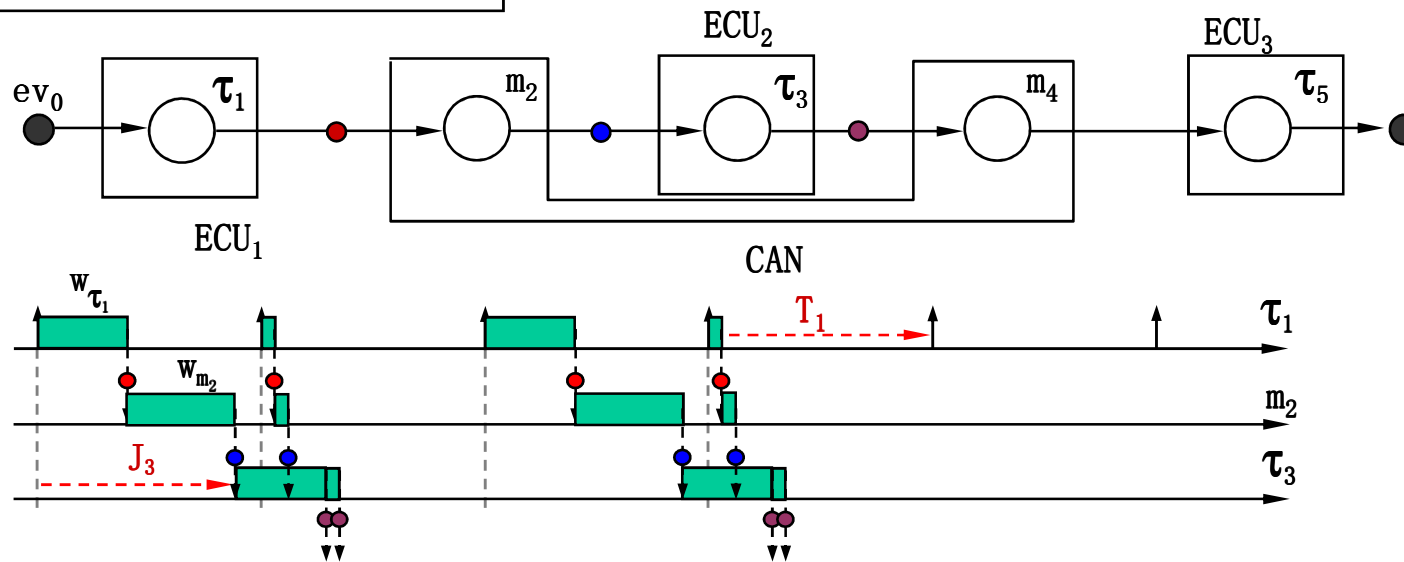
$$r_i = C_i + \sum_{j \in hp(i)} \left\lceil \frac{r_i}{T_j} \right\rceil C_j$$

# Event-based Activation Model

Lower latency for high priority paths, jitter increases along the path

End-to-end latency analysis

Data-driven precedence constrained activation model

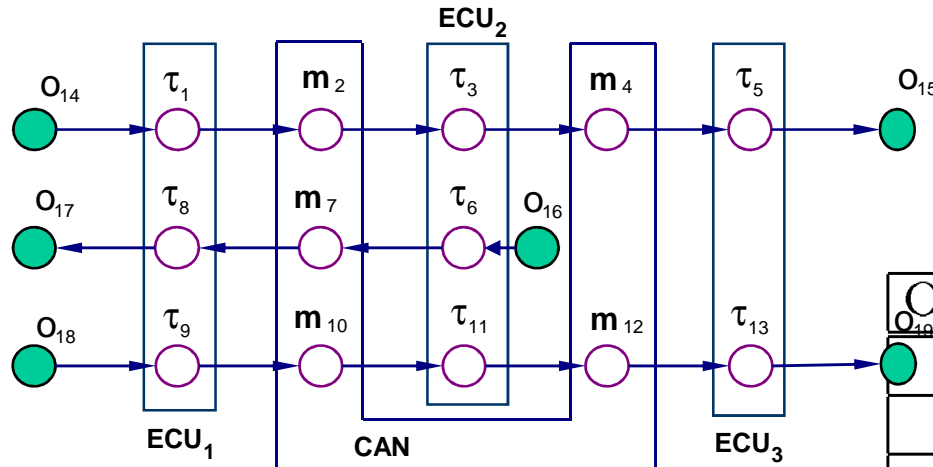


$$l_{(i,j)} = \sum_{k: o_k \in P(i,j)} w_k$$

where (approx.)

$$w_i = C_i + \sum_{j \in hp(i)} \left\lceil \frac{w_i + J_j}{T_j} \right\rceil C_j$$

# Activation modes: latency tradeoffs



## End to end latency requirements

$$d_{O_{14}, O_{15}} \Rightarrow \mathbf{70}$$

$$d_{O_{16}, O_{17}} \Rightarrow \mathbf{100}$$

$$d_{O_{18}, O_{19}} \Rightarrow \mathbf{120}$$

## Mixed activation mode

$$L_{i,j} = \sum_{k:k=j \vee l_{k,q} \in \mathcal{E}_p} (J_k + w_k) + \sum_{q:l_{k,q} \in \mathcal{E}_p} T_q$$

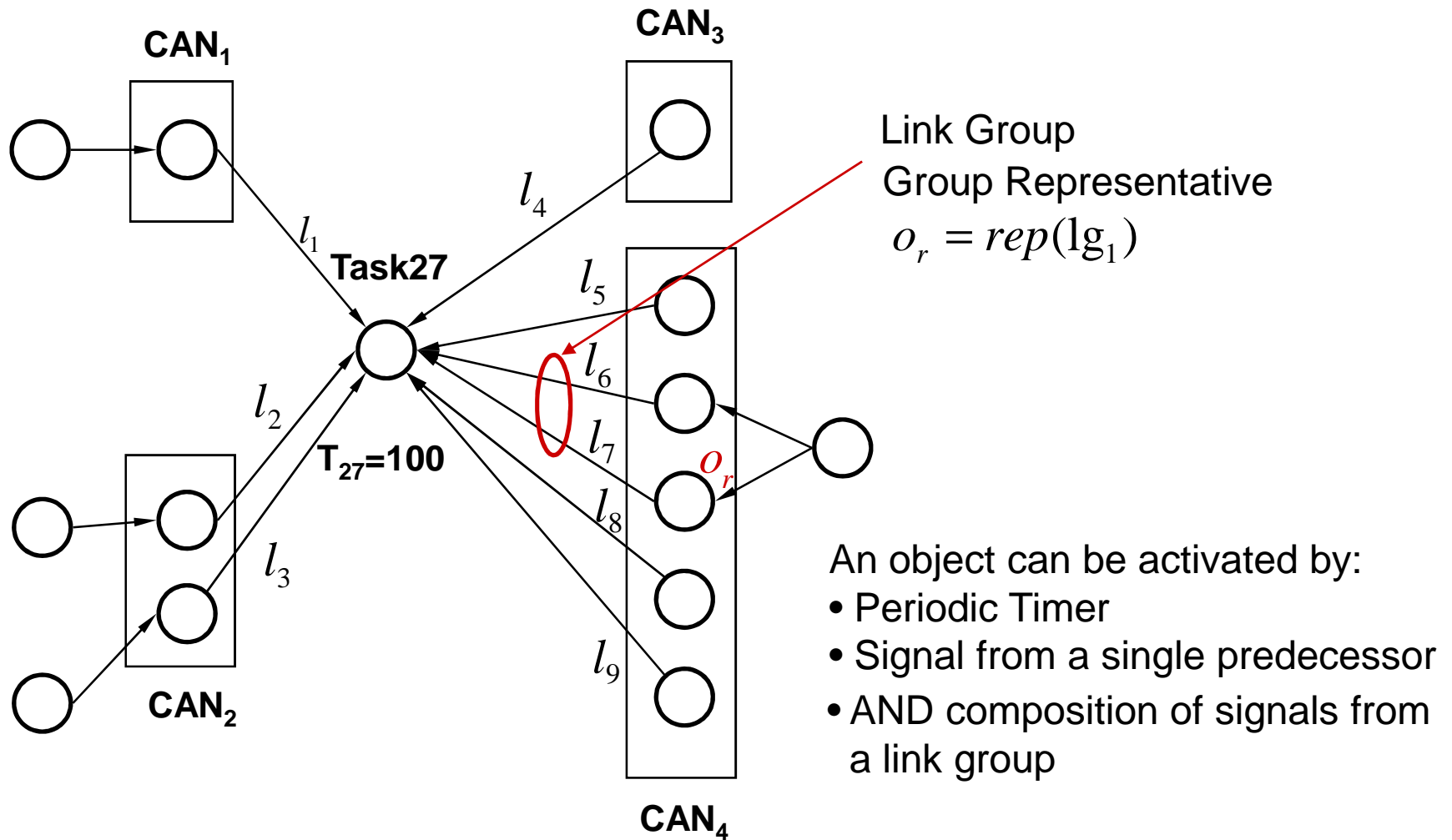
Periodic

Event-based

Object	$\pi_i$	$T_i$	$C_i$	$r_i$	$l_i$	$J_i$	$w_i$	$r_i$
$\tau_1$	13	15	8	8	8	0	8	8
$m_2$	12	15	2	4	27	8	4	12
$\tau_3$	11	15	8	8	50	12	8	20
$m_4$	10	15	2	6	71	20	6	26
$\tau_5$	9	15	6	6	<del>92</del>	26	6	<b>32</b>
$\tau_6$	8	40	6	14	14	0	30	30
$m_7$	7	40	2	8	62	30	12	42
$\tau_8$	6	40	14	30	<del>122</del>	42	30	<b>72</b>
$\tau_9$	5	30	2	42	42	0	190	190
$m_{10}$	4	30	2	10	82	190	18	208
$\tau_{11}$	3	30	6	28	140	208	58	266
$m_{12}$	2	30	2	10	180	266	36	302
$\tau_{13}$	1	30	8	14	<del>224</del>	302	32	<del>324</del>

# Model Definition

- Selection of the activation event and link groups

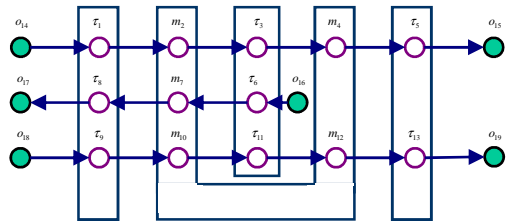


# Latencies of OSEK Tasks and CAN Messages

		Linear Combination	First Instance
Processor	Upper	$w_i^\uparrow = C_i + \sum_{j \in hp(i)} \left( \frac{w_i^\uparrow + J_j}{T_j} + 1 \right) C_j$ $w_i(q) = (q+1)C_i + \sum_{j \in hp(i)} \left( \frac{w_i(q) + J_j}{T_j} + 1 \right) C_j$	$w_i = C_i + \sum_{j \in hp(i)} \left\lceil \frac{w_i + J_j}{T_j} \right\rceil C_j$
	Lower	$w_i = \max_q \{ w_i(q) - qT_i \}$ $r_i = J_i + w_i$ $q = 0 \dots q^*, r_i(q^*) \leq T_i$	$r_i = J_i + w_i$
Bus	Upper	$wq_i^\uparrow = B_i + \sum_{j \in hp(i)} \left( \frac{wq_i^\uparrow + J_j}{T_j} + 1 \right) C_j$ $wq_i(q) = (q+1)B_i + \sum_{j \in hp(i)} \left( \frac{wq_i(q) + J_j}{T_j} + 1 \right) C_j$	$wq_i = B_i + \sum_{j \in hp(i)} \left\lceil \frac{wq_i + J_j}{T_j} \right\rceil C_j$
	Lower	$wq_i = \max_q \{ wq_i(q) - qT_i \}$ $r_i = J_i + wq_i$ $q = 0 \dots q^*, r_i(q^*) \leq T_i$	$w_i = wq_i + C_i$ $r_i = J_i + w_i$

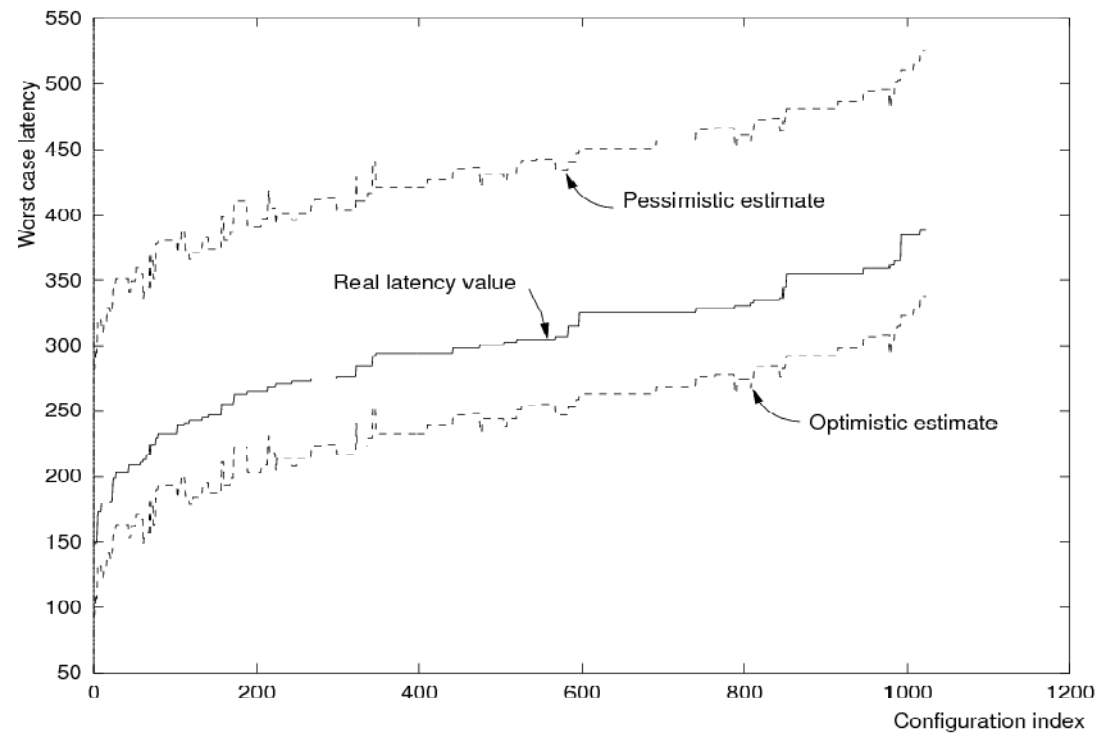


# Linear Approximation



	$L_{o_{14},o_{15}}$	$L_{o_{16},o_{17}}$	$L_{o_{18},o_{19}}$
<i>Linear_upper</i>	44.36	130.86	507.03
<i>Fixed_point</i>	40	88	312
<i>Linear_lower</i>	38.91	79.43	294.96

A linear combination of linear upper and lower bounds can be sufficiently accurate to be used as an estimator of actual e2e latency



# MILP Solution

## Sets

$V$ : Set of objects implementing the computation and communication functions  
 $E$ : Set of links connecting schedulable objects  
 $R$ : Set of resources (CAN, ECUs)

## Parameters

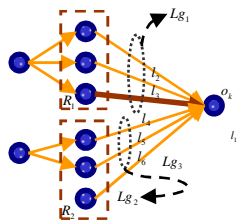
$\pi_i$ : Priority of object  $o_i$   
 $T_i$ : Period of object  $o_i$   
 $C_i$ : Worstcase execution/transmission time of object  $o_i$

## Variables

$r_i$ : Worst case response time of object  $o_i$   
 $J_i$ : Release Jitter of object  $o_i$   
 $w_i$ : Worstcase runnable queueing time of object  $o_i$   
 $L_{s,t}$ : End to end latency between object  $o_s$  and  $o_t$   
 $y_{h,k} = \begin{cases} 1, & \text{If activation of } o_k \text{ is event-driven by } o_h \\ 0, & \text{otherwise} \end{cases}$

# Feasibility Constraints 1

## Jitter Inheritance Rule



$$y_{r,k} = y_{s,k}$$

$$\sum_{Lg_h \in G(o_k)} y_{r,k} \leq 1$$

$$J_k \leq \sum_{Lg_h \in G(o_k)} y_{r,k} \times M$$

$$0 \leq J_k$$

$$J_k \leq r_r + (1 - y_{r,k}) \times M$$

$$r_r - (1 - y_{r,k}) \times M \leq J_k$$

$$r_h + (y_{h,k} - 1) \times M \leq J_k$$

$$J_k \leq r_h$$

$$J_k \leq y_{h,k} \times M$$

All links in one group assume the same activation model

Only one of the incoming link group can provide its activation signal

If none of incoming groups carry activation signal, then release jitter of object k is 0

Release jitter inherited from object r which has largest wcr from the activating group

Simplified version of link groups

# Feasibility Constraints 2

## WCRT Rule

$$w_h = r_h \bar{C}_h + \sum_{k \in hp(h)} \left( \frac{w_h + J_k}{T_k} + \alpha \right) C_k$$

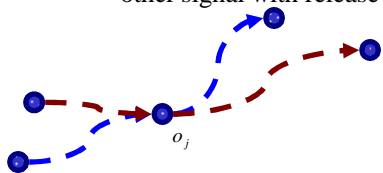
$$\sum_{P_r \in P} (\alpha \times L_{P_r}^\uparrow + (1 - \alpha) \times L_{P_r}^\downarrow - L_{P_r})^2$$

Calculation of worst case response time

- A linear combination of linear upper and lower bounds is used as an estimation of runnable queuing time
- alpha is chosen to minimize the mean square fit function

## Latency Rule

$$z_{i,j} = \begin{cases} w_j & \text{if link } l_{i,j} \text{ carries activation signal} \\ w_j + J_j + T_j & \text{otherwise, } o_j \text{ may be activated by} \\ & \text{other signal with release jitter } J_j \end{cases}$$



Path end to end latency can not exceed deadline

$$w_j \leq z_{i,j}$$

$$z_{i,j} \leq w_j + (1 - y_{i,j}) \times M$$

$$z_{i,j} \leq w_j + J_j + T_j$$

$$w_j + J_j + T_j - y_{i,j} \times M \leq z_{i,j}$$

$$L_{s,t} = \sum_{l_{u,v} \in P_{s,t}} z_{u,v}$$

$$L_{s,t} \leq d_{s,t}$$

# Possible Objective Function

$$\text{Maximize } \sum_{Lg_h \in G} y_{j,k}$$

Minimization of the number of event buffers in the system

$$\text{Minimize } \sum_{P_r \in P} L_{p_r}$$

Minimization of sum of end to end latencies

$$\text{Minimize } \sum_{P_r \in P} \gamma_{p_r} \times \text{Max}(L_{p_r} - d_{p_r}, 0)$$

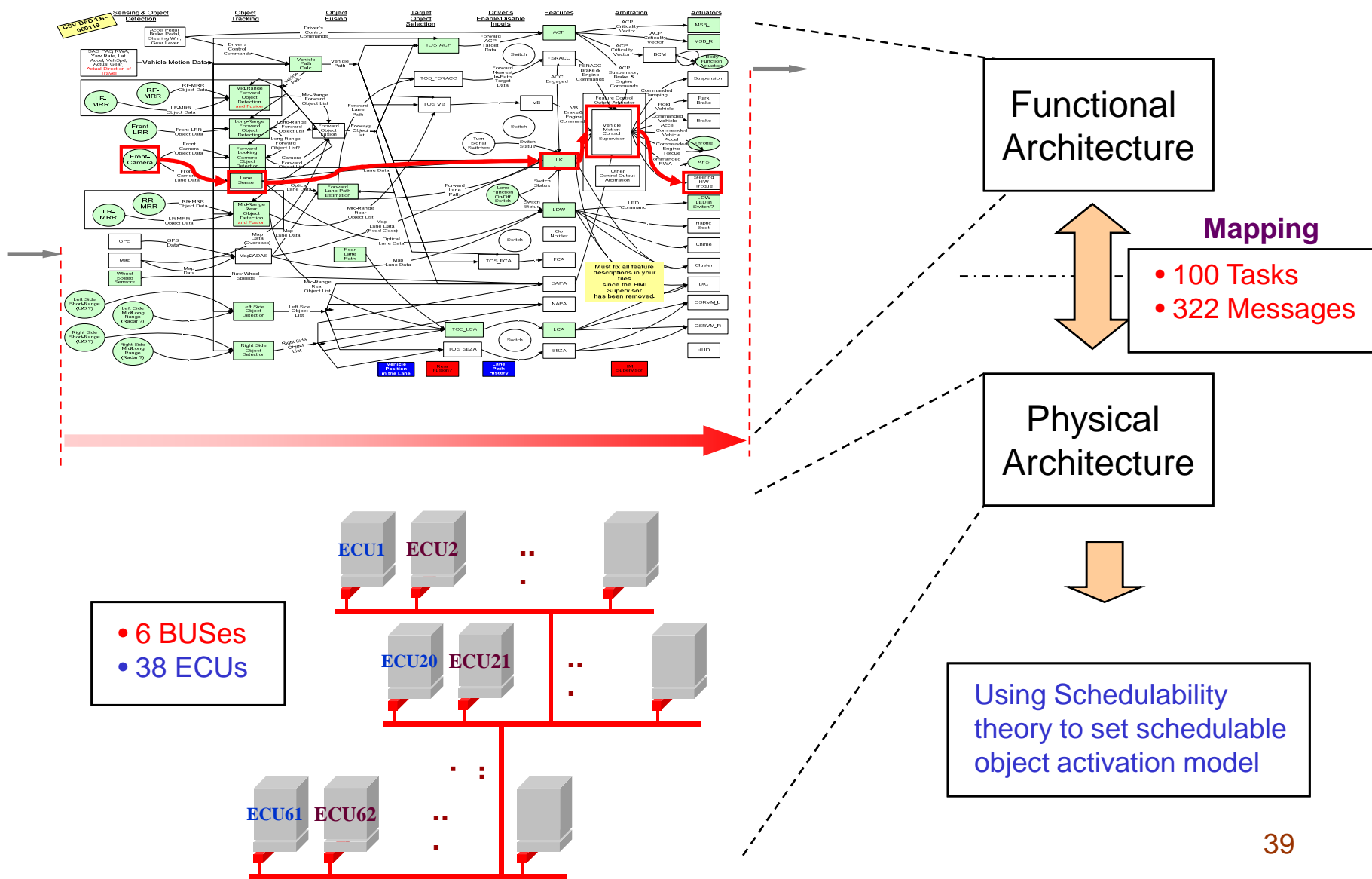
Minimization of sum of weighted deadline violation



## Outline

- Automotive architecture trends and challenges
- System-level design methodology and timing evaluation
- Activation models and end-to-end latencies
- Problem definition
  - Example
- MILP Optimization
- **Case Study**

# Experimental vehicle case study



# Case study results

## Before Optimization (all periodic)

- Worst case = **577ms** was found for paths with deadline **300ms**
- Worst case = **255.5ms** found for paths with deadline **200ms**
- Worst case = **145.4ms** found for paths with deadline **100ms**

## Problem characterization

- **38 ECUs, 6 Buses**
- **Bus speed between 25 and 500 kb/s**
- **Bus utilization between 30% to 50%**
- **CPU utilization between 5% to 60%**
- **100 tasks, 322 messages**
- **Number of links in the functional dataflow is 507**
- **184 Paths analyzed between 10 pairs of functional nodes**

## Optimization results

- A feasible solution is found if using the largest lateness path metric after changing 24 groups
- **294.8** for paths with  $d=300$
- **158.1** for paths with  $d=200$
- **95.46** for paths with  $d=100$  (**61.57 average slack**)
- the solution was improved with 5 extra branches (**76.79 average slack**)
- $\alpha$  practically constant = **0.465** with weighted sum of path latencies (evaluating all nodes) no solution found

## Time to solve is

- **2.6 s** for the exact analysis
- **7 s** for the linear approx (on a 1.4GHz PC)



# Approach

- Mathematical programming
  - Modifying an object period affects multiple paths
  - Additional constraints due to legacy tasks and messages
- Geometric Programming: Poly-time optimization

- Standard Form:

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 1 \quad i = 1, \dots, m \\ & && g_i(x) = 1 \quad i = 1, \dots, p \end{aligned}$$

- $x = (x_1, x_2, \dots, x_n)$  are positive real-valued variables
- $g$  is a set of monomial functions

$$m(x) = cx_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \quad c > 0, a_i \in \mathbb{R}$$

- $f$  is a set of posynomial functions
  - Sum of monomials

# Geometric programming formulation

- Approximate the response time  $r_i$  with  $s_i$ 
  - $0 \leq \alpha_i \leq 1$
  - If all  $\alpha_i = 1$ ,  $s_i \geq r_i$

$$s_i = c_i + \sum_{j \in hp(i)} \left( \frac{s_i}{t_j} + \alpha_i \right) c_j \quad \forall o_i \in \mathcal{T}$$

**Minimize the sum of approx. response times**  $\rightarrow$  *min.*  $\sum_{o_i \in \mathcal{O}} s_i$

**Meet end-to-end latency deadlines**  $\rightarrow$  *s.t.*  $\ell_p \leq d_p \quad \forall p \in \mathcal{P}$

**Transformed equations for approx. response times**  $\left\{ \begin{array}{l} \frac{\sum_{j \in hp(i)} c_j \alpha_i + c_i}{s_i} + \sum_{j \in hp(i)} \frac{c_j}{t_j} \leq 1 \quad \forall o_i \in \mathcal{T} \\ \frac{\sum_{j \in hp(i)} c_j \alpha_i + b_i}{s'_i} + \sum_{j \in hp(i)} \frac{c_j}{t_j} \leq 1 \quad \forall o_i \in \mathcal{M} \\ s_i = s'_i + c_i \quad \forall o_i \in \mathcal{M} \end{array} \right.$

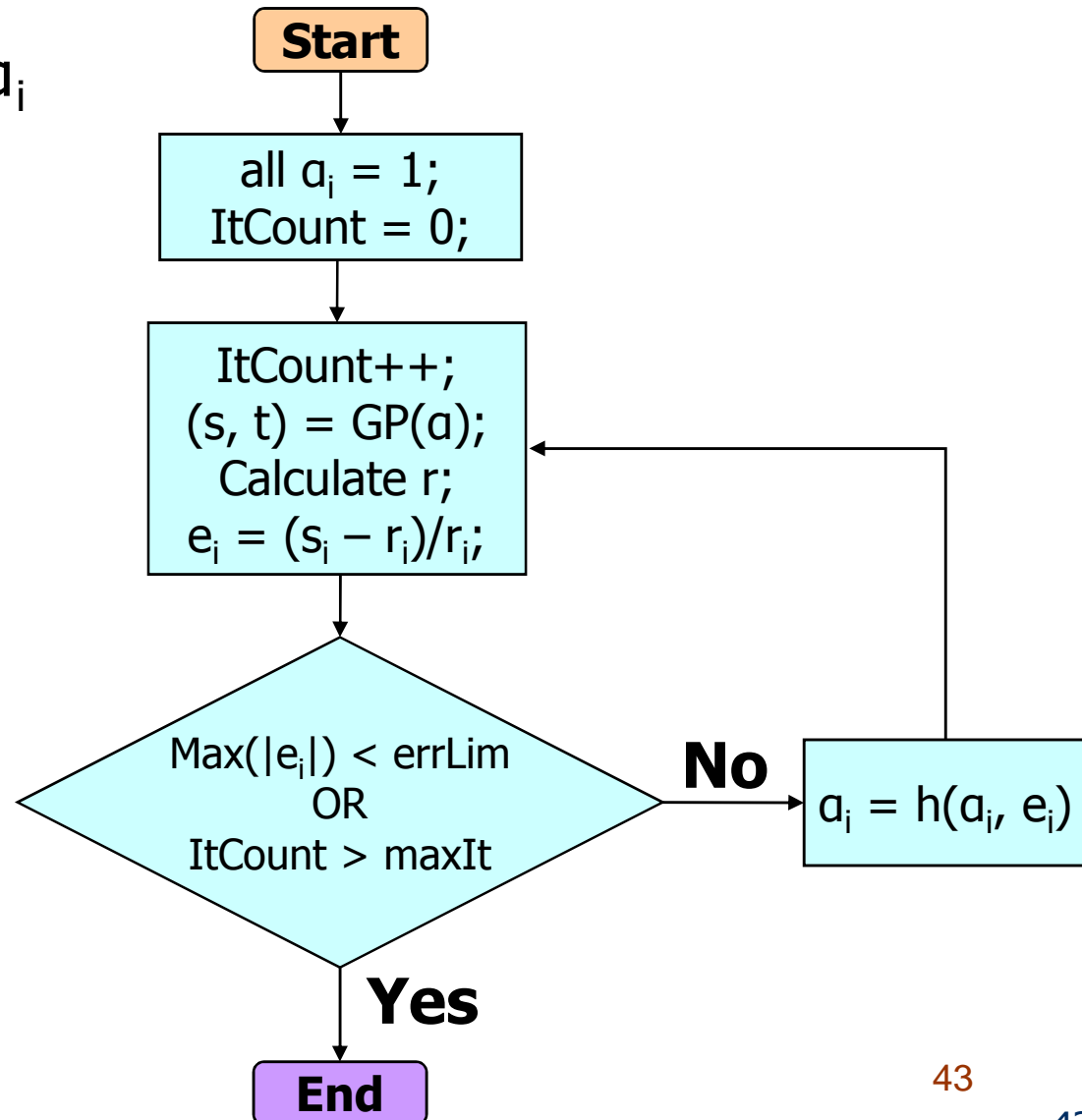
**Ensure schedulability**  $\rightarrow$   $\frac{s_i}{t_i} \leq 1 \quad \forall o_i \in \mathcal{O}$

**Meet utilization bounds**  $\rightarrow$   $\sum_{o_i | R_{o_i} = j} \frac{c_i}{t_i} \leq u_j \quad \forall R_j \in \mathcal{R}$

**Lower and upper bounds for periods**  $\rightarrow$   $n_i \leq t_i \leq x_i \quad \forall o_i \in \mathcal{O}$

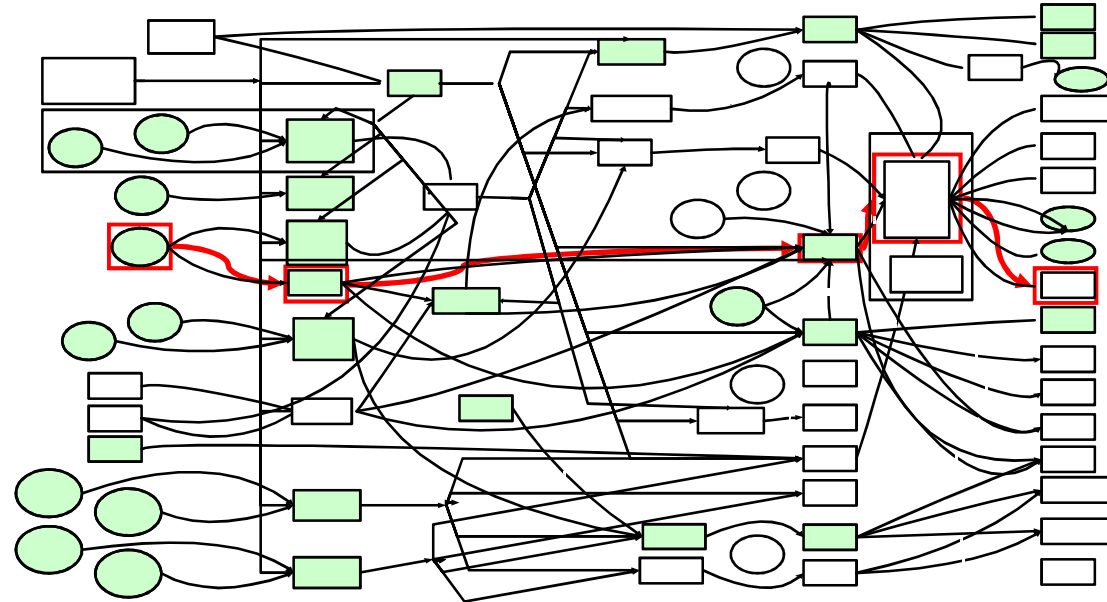
# Iterative Procedure to Reduce Error

- Iteratively change  $a_i$  based on error
- Parameters
  - maxIt - max. # of iterations
  - errLim - max. permissible error



## Case Study: Advanced Safety Vehicle

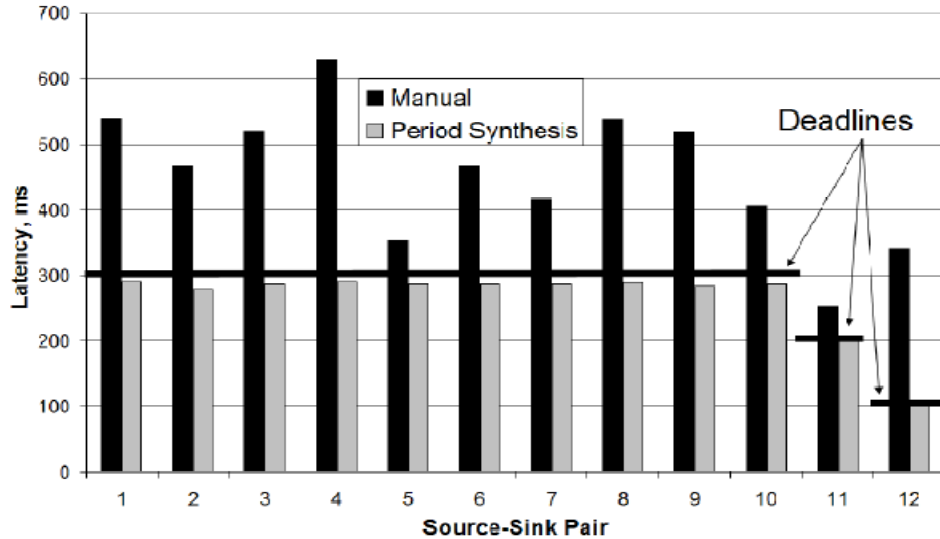
- From GM Research
- E.g. enhanced cruise control, lane departure warning, parallel parking assist
- Architecture
  - 38 ECUs
  - 4 buses
- Functionality
  - 92 tasks
  - 196 messages



- End-to-end latency constraints
  - Over 12 source-sink task pairs
  - 222 total paths
  - Deadlines range from 100ms to 300ms

# Experiments

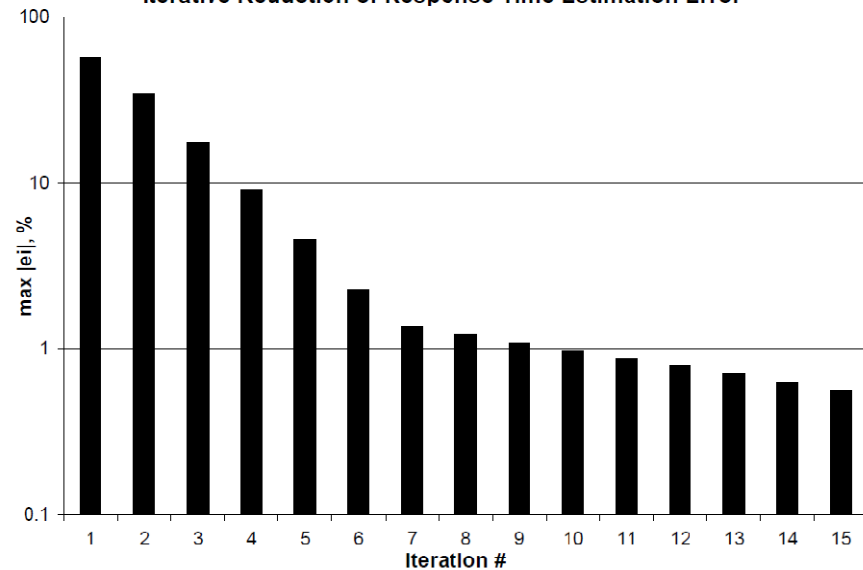
Latency Before and After Period Synthesis



- GP optimization meets all deadlines in 1<sup>st</sup> iteration
- Solution time: 24s

- Maximum error reduced from 58% to 0.56% in 15 iterations
- Average error (not shown) reduced from 6.98% to 0.009%

Iterative Reduction of Response Time Estimation Error





## Concluding remarks

- Quantitative analysis offers opportunities for architecture exploration and selection
- Domains of cost, dependability and time have been identified as prime candidates
  - not considering, for example, power
- Analysis techniques are at different levels of maturity
- Uncertainty challenge
  - Some required information is typically not available in the early development stages
  - Requirements extraction process is not mature
- Synthesis to be extended to other domains
  - leveraging MILP or GP formulations of the placement, priority assignment and period definition problems



## Concluding remarks

- Worst case timing analysis can be applied to design optimization problems
- With respect to end-to-end latencies in distributed architectures there are multiple dimensions that can be explored
  - task allocation
  - period assignment
  - priority assignment
  - ...
- Also, most active safety functions are not truly hard real-time and worst case analysis may be pessimistic
  - end-to-end stochastic analysis
  - design optimizations based on stochastic analysis <sup>2</sup><sub>47</sub>



## Q&A

Thank you!







E-mail: [marco@sssup.it](mailto:marco@sssup.it)