Carnegie Mellon
Electrical & Computer ENGINEERING

On Understanding Emergence In the Context of System Safety

Philip Koopman, Jen Black

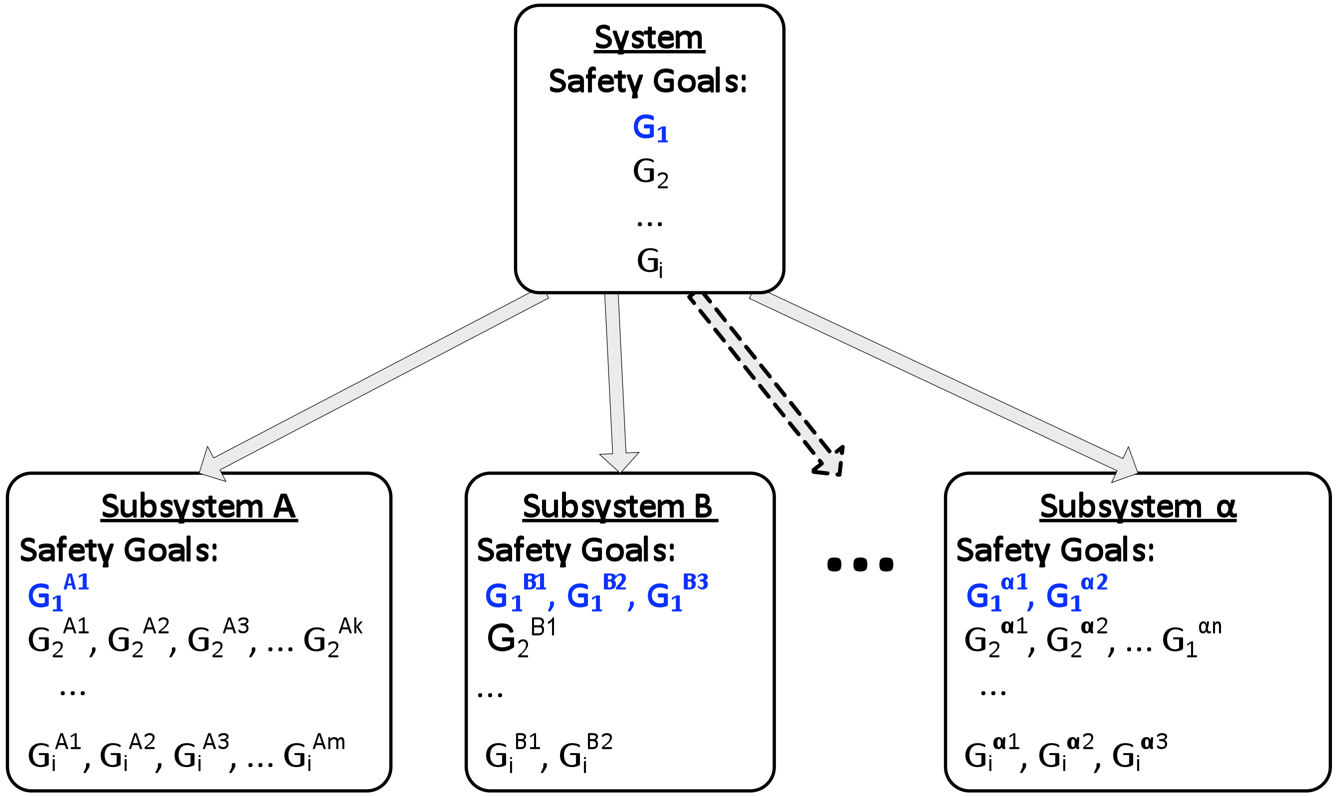
(Based on thesis work of Jen Black – DSN 2009 paper)

IFIP WG 10.4 Meeting, Obidos Portugal
July 5th, 2009

Motivation

- Want to decompose safety critical functions to subsystems
  - Increases chance of system integration success
  - Permits testing for safety before system integration
    - Example: vehicle has subsystems from many suppliers
- But, safety is emergent
  - Loosely: examination of individual components doesn't completely predict safety
- Research topic (Jen Black DSN 2009 paper & thesis):
  - Decomposition of emergent properties is undecidable
  - Can we do something useful in the non-worst-case?

How do We Decompose Safety?

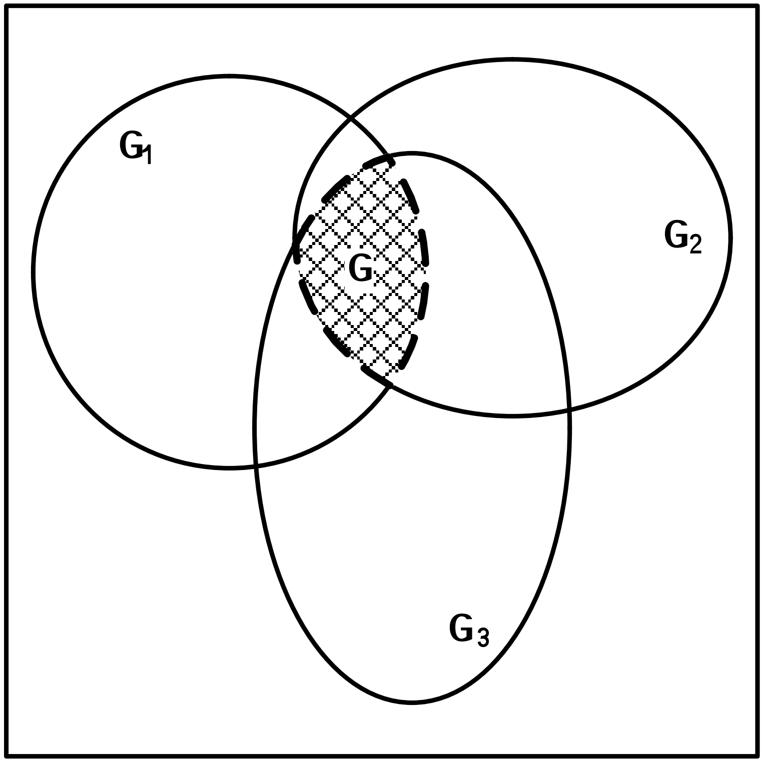


- $G_1^{A1} \wedge G_1^{B1} \wedge G_1^{B2} \wedge G_1^{B3} \wedge \ldots G_1^{\alpha 1} \wedge G_1^{\alpha 2} \Leftrightarrow G$
  - Can this be done? Where does emergence fit in?
  - Is partial decomposition possible? Is it useful?

Definitions: Fully Composable, Emergent

- G is **fully composable** if:
  $\exists \{G_1, G_2, \ldots G_n\}$ such that:
  $G_1 \wedge G_2 \ldots \wedge G_n \Leftrightarrow G$

  which can also be expressed as:
  $(G_1 \wedge G_2 \ldots \wedge G_n \Rightarrow G)$
  $\wedge (\neg G_1 \vee \neg G_2 \ldots \vee \neg G_n \Rightarrow \neg G)$



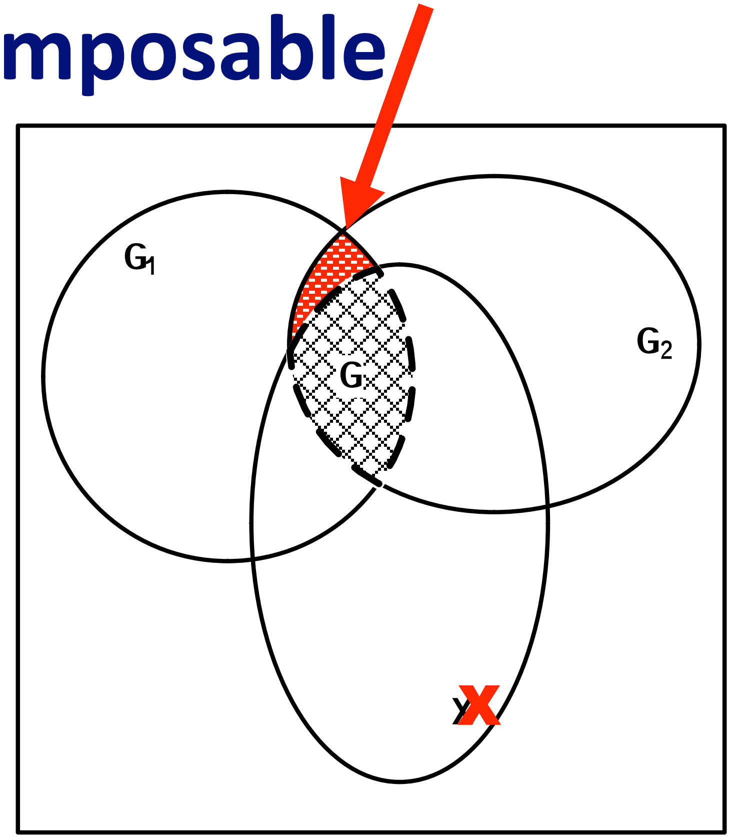
Example goal: *ObjectInPath ⇒ StopVehicle*

Subgoals: *(ObjectInPath ⇔ CA.StopVehicle) ∧ (CA.StopVehicle ⇒ StopVehicle)*

detector subsystem **brake subsystem**

- G is **Emergent** if no such set of subgoals exists

Emergent but Partially Composable



- **G is <u>partially composable</u> if:**

  $\exists \{G_1, G_2, \ldots G_m\}$, **X** *(emergence)*

  such that:

  $G_1 \wedge G_2 \ldots \wedge G_m \wedge X \Leftrightarrow G$

which can also be expressed as:

$(G_1 \wedge G_2 \ldots \wedge G_m \wedge X \Rightarrow G)$

$\wedge (\neg G_1 \vee \neg G_2 \ldots \vee \neg G_m \vee \neg X \Rightarrow \neg G)$

- **<u>Composability</u>** is the degree to which **X** is small

The Key Idea

- Functional correctness is about doing the ***right*** thing:

  $(G_1 \wedge G_2 \ldots \wedge G_n \wedge \mathbf{X} \Rightarrow G)$

  - Over-approximating any G_i is OK
  - If you are missing any sub-goal **X**, you don't achieve G

- BUT, safety is often about ***not*** doing the ***wrong*** thing:

  $(\neg G_1 \vee \neg G_2 \ldots \vee \neg G_n \wedge \neg \mathbf{X} \Rightarrow \neg G)$

  - Identifying ***any*** of the subgoals can be useful
  - Under-approximating ¬ G still increases safety, even without **X**

- Results: Identifying only some subgoals seems helpful
  - Monitoring sub-goals at run-time; process of finding subgoals too
  - Found 11 safety-critical design defects on research vehicle model