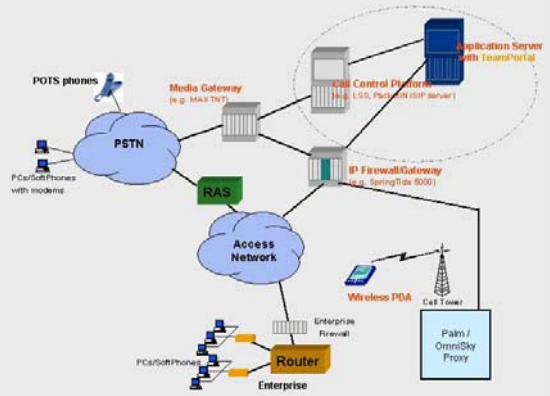
Research Methods and Theory in Global Software Development

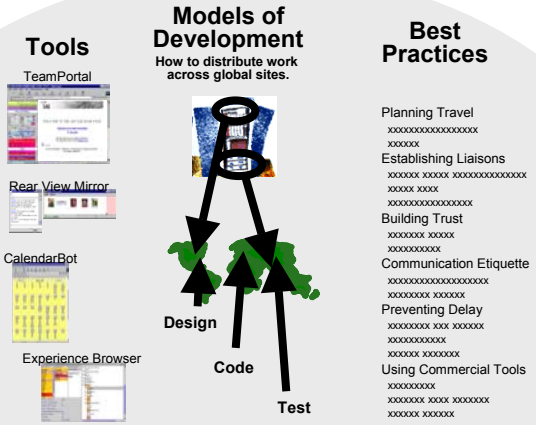
Jim Herbsleb
School of Computer Science
Carnegie Mellon University

Bell Labs Collaboratory

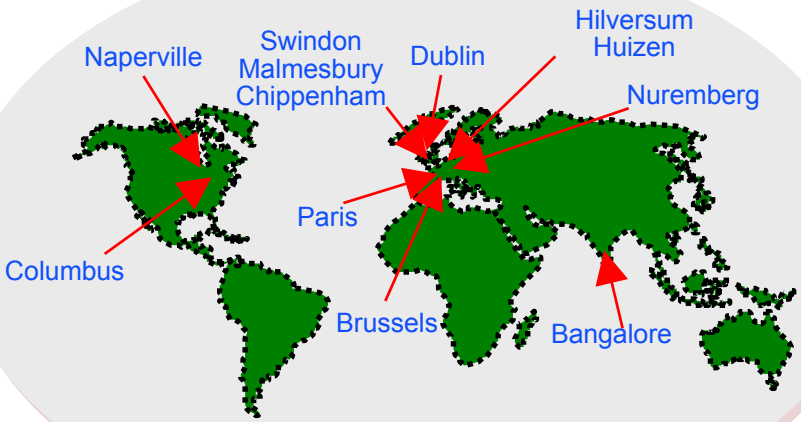


New Products

Research Team

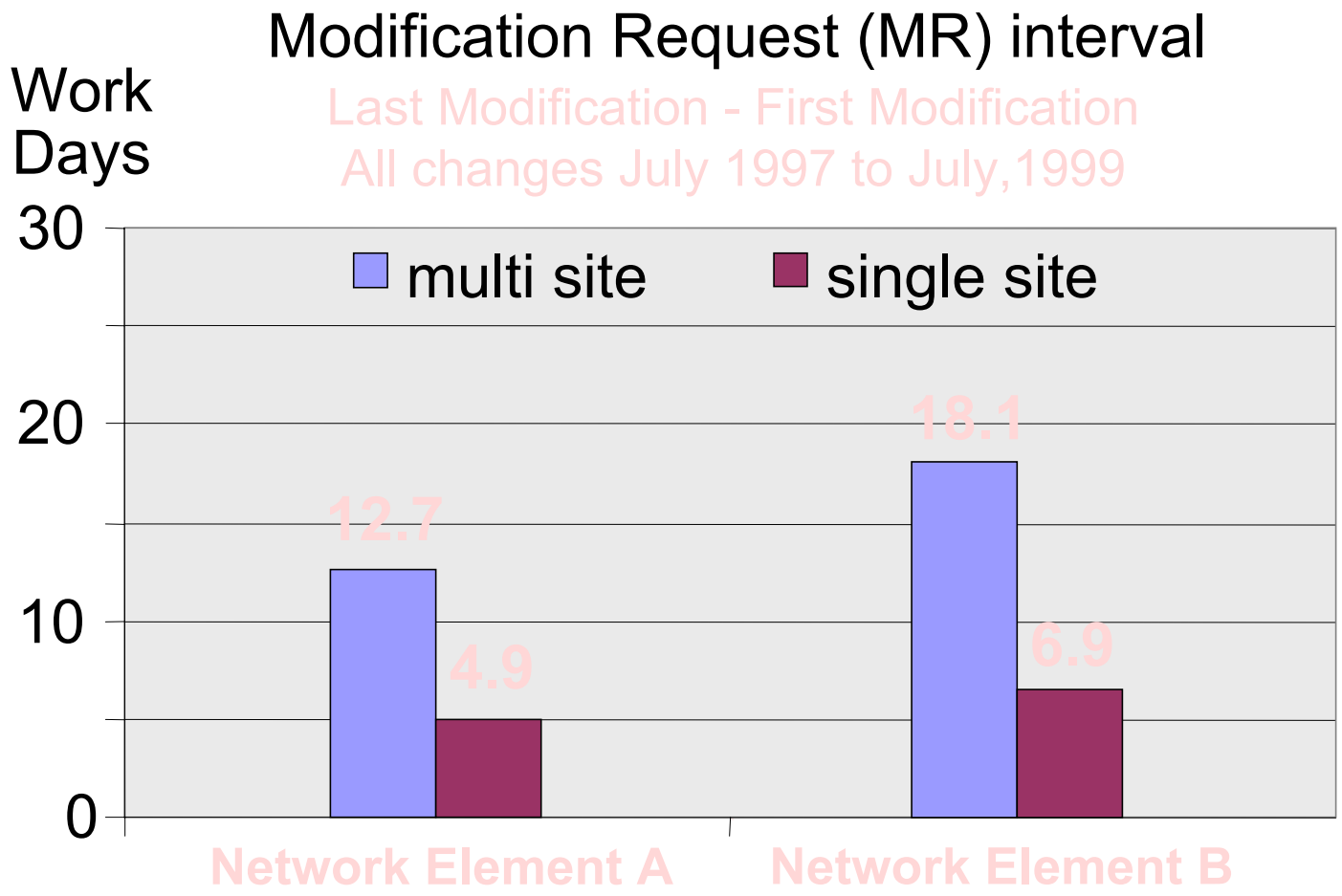


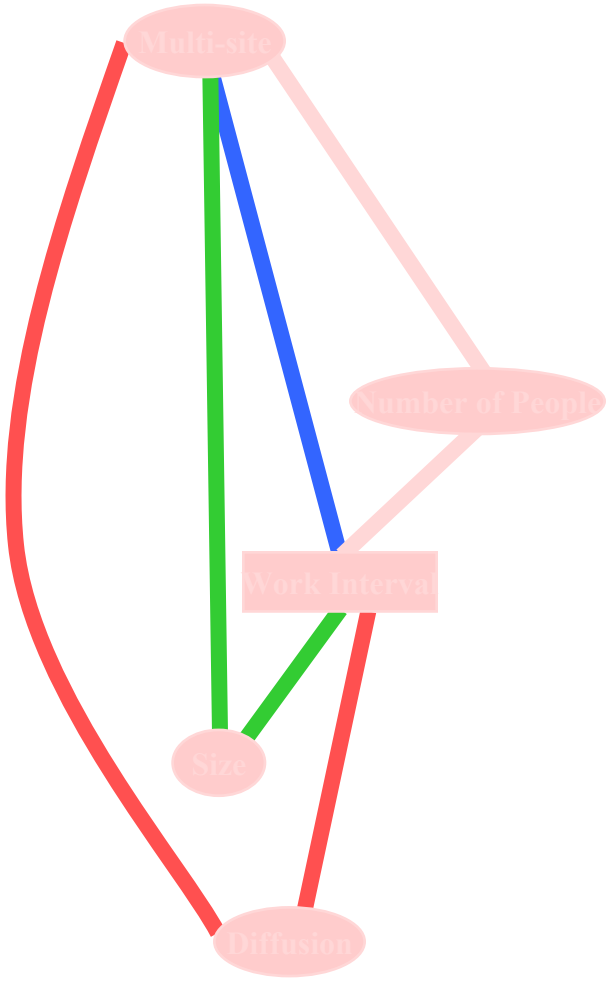
Global Development Solutions



Empirical Studies

Multi-site Delay





H1 Multi-site work just takes longer

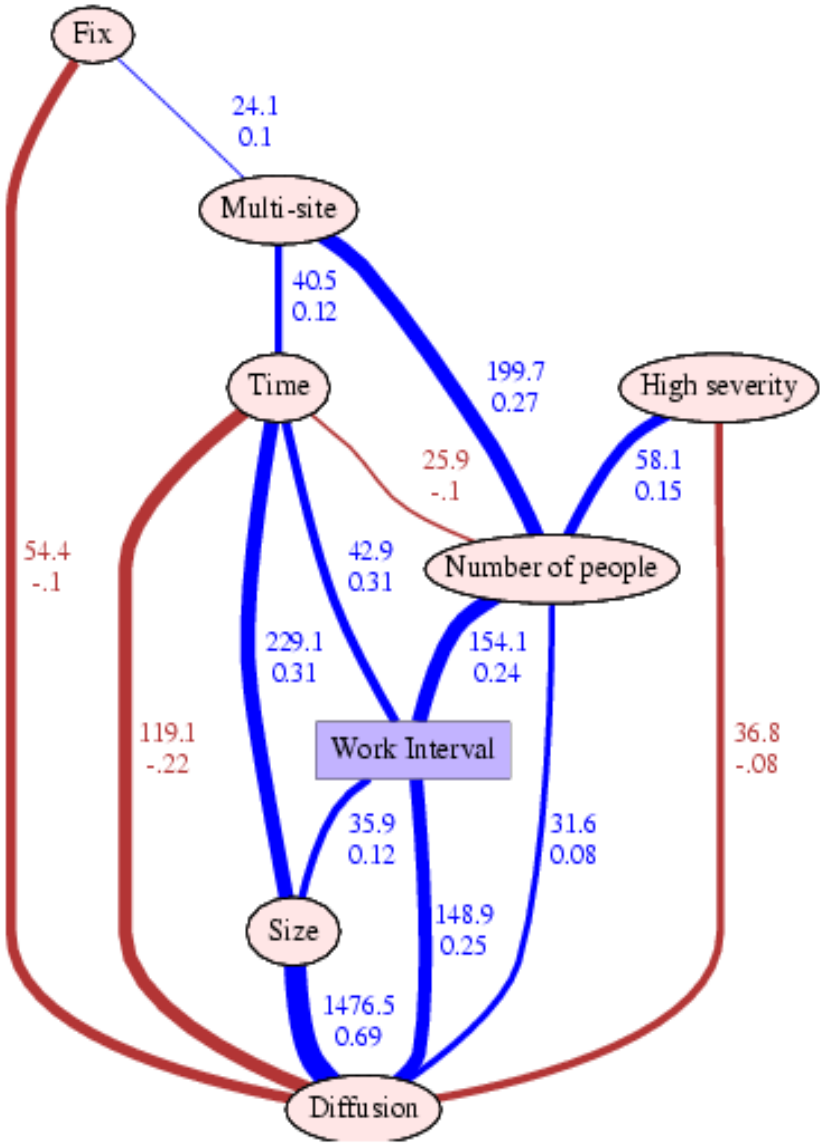
H2 Multi-site MRs are larger, take longer

H3 Multi-site MRs are more diffuse, take longer

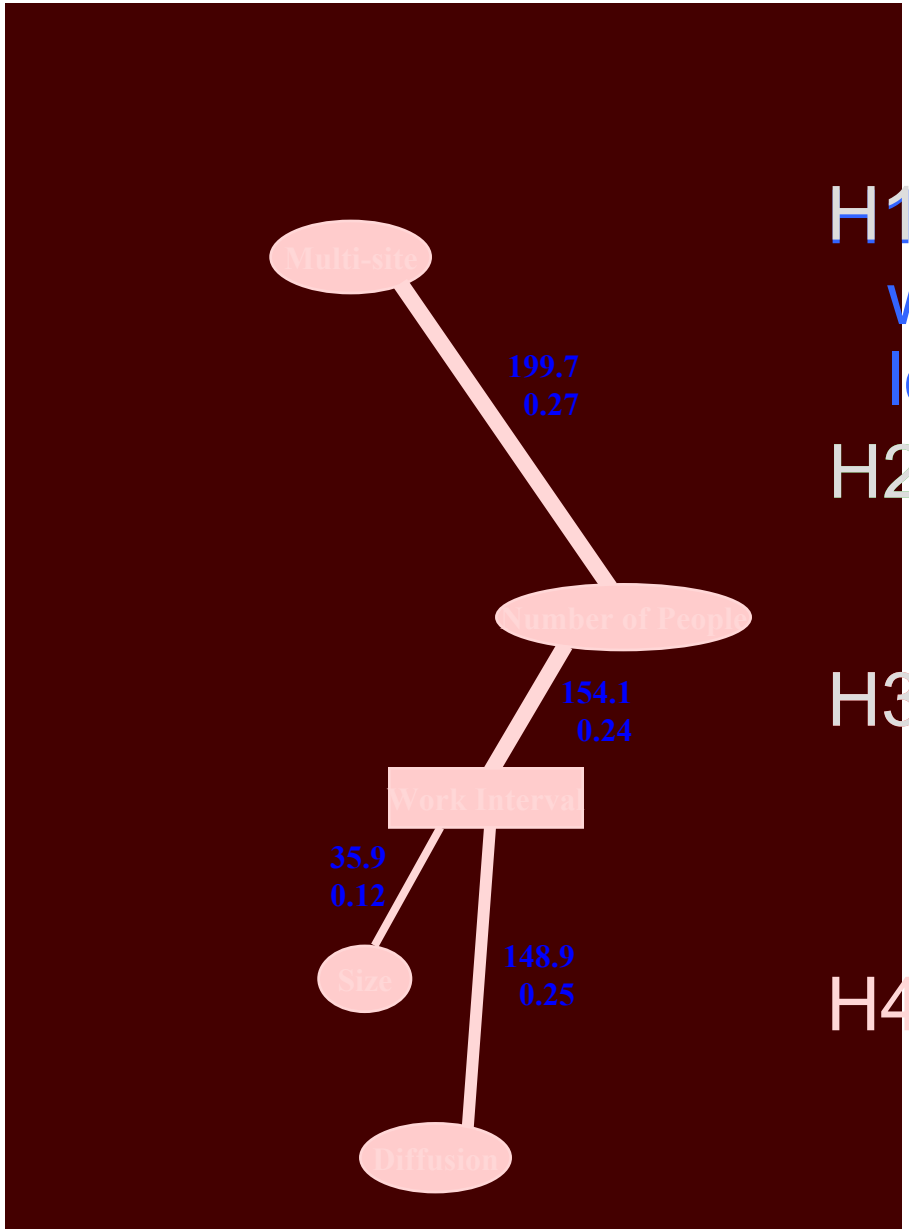
H4 Multi-site MRs involve more people, take longer

Modeling Interval

Variable	Measure used in models
MR interval	Log of number of days, first delta to last delta
Number of people	Log of number of people
Diffusion	Log of number of modules touched by change
Size	Log of number of delta
Time	Date
Severity	Is high severity
Fix	Is fix
Multi-site	Set of sites of all actors has more than one element



Graphical model of work interval for Network Element A



H1 Multi-site work just takes longer

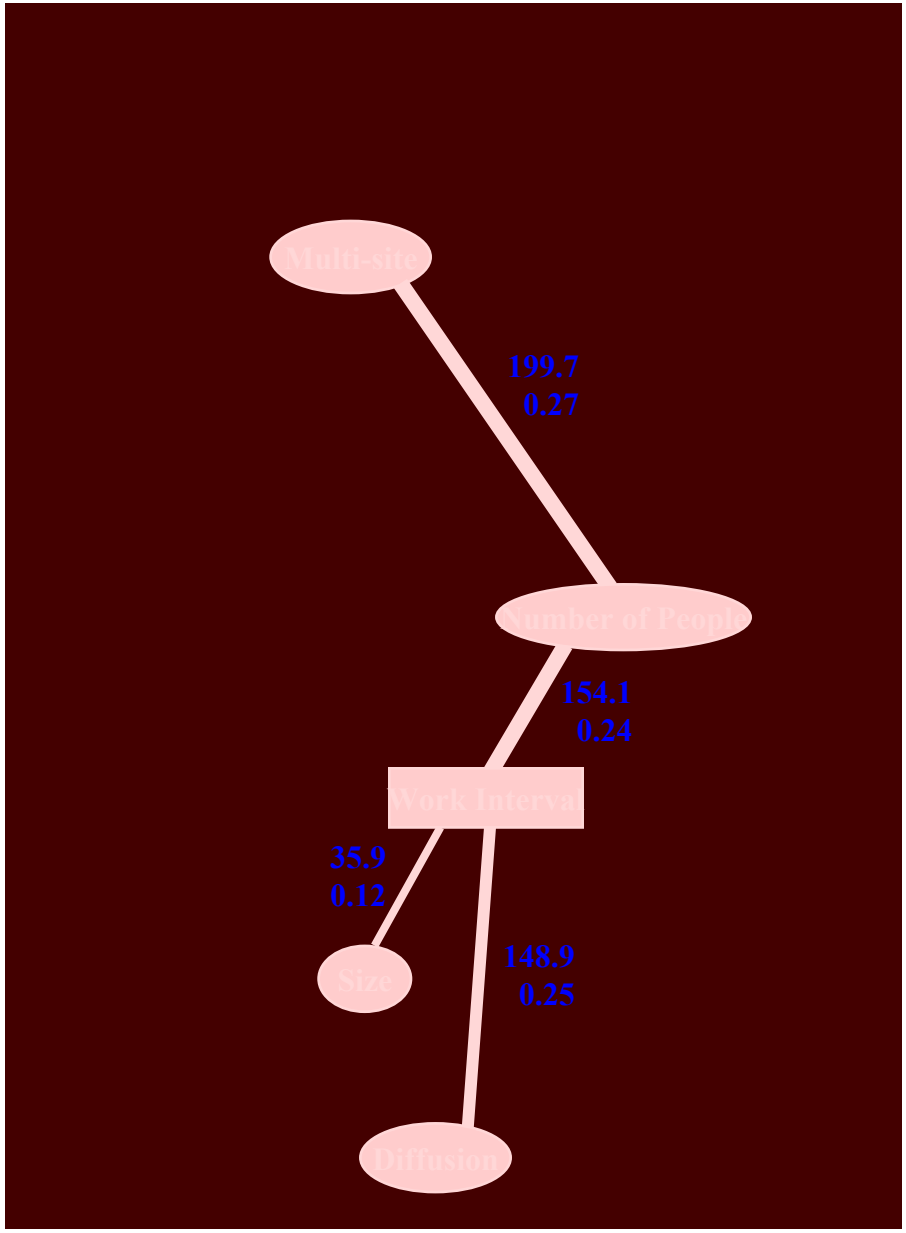
Multi-site takes longer

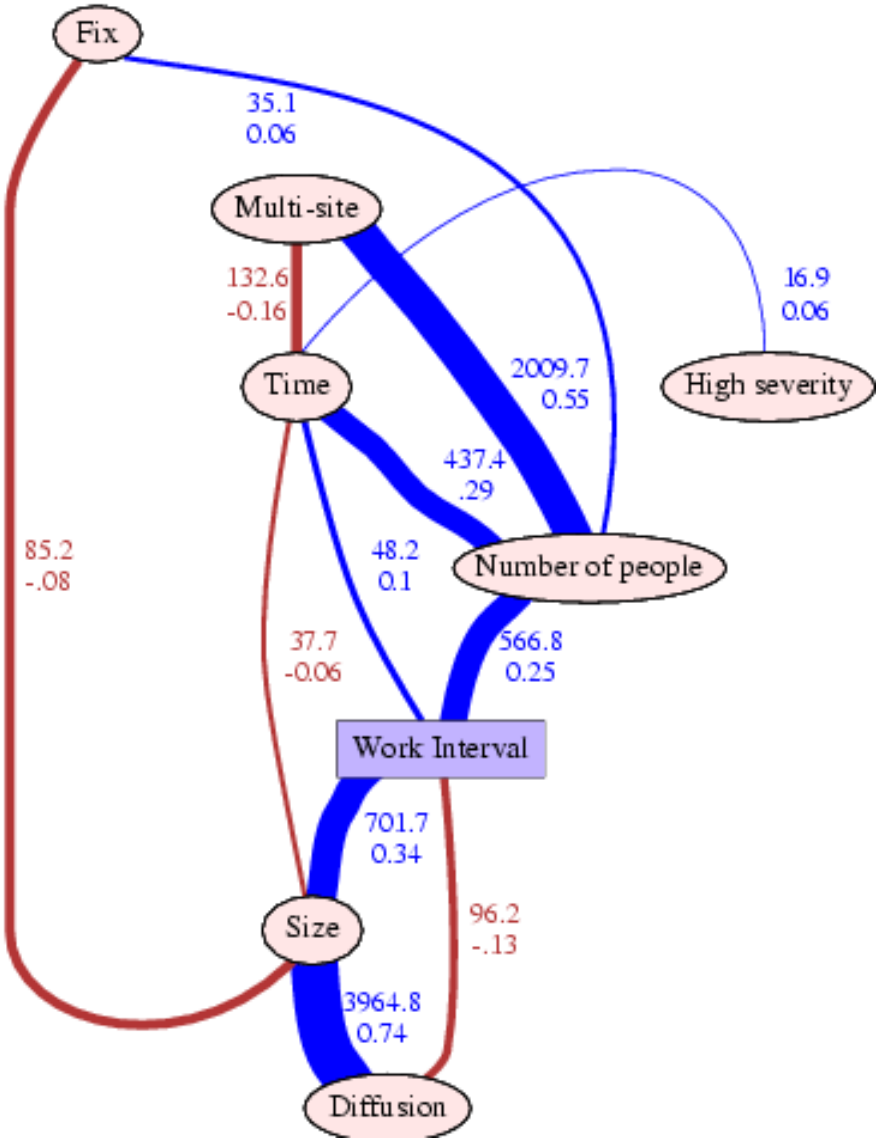
H2 Multi-site MRs are larger, take longer

H3 Multi-site MRs are more diffuse, take longer

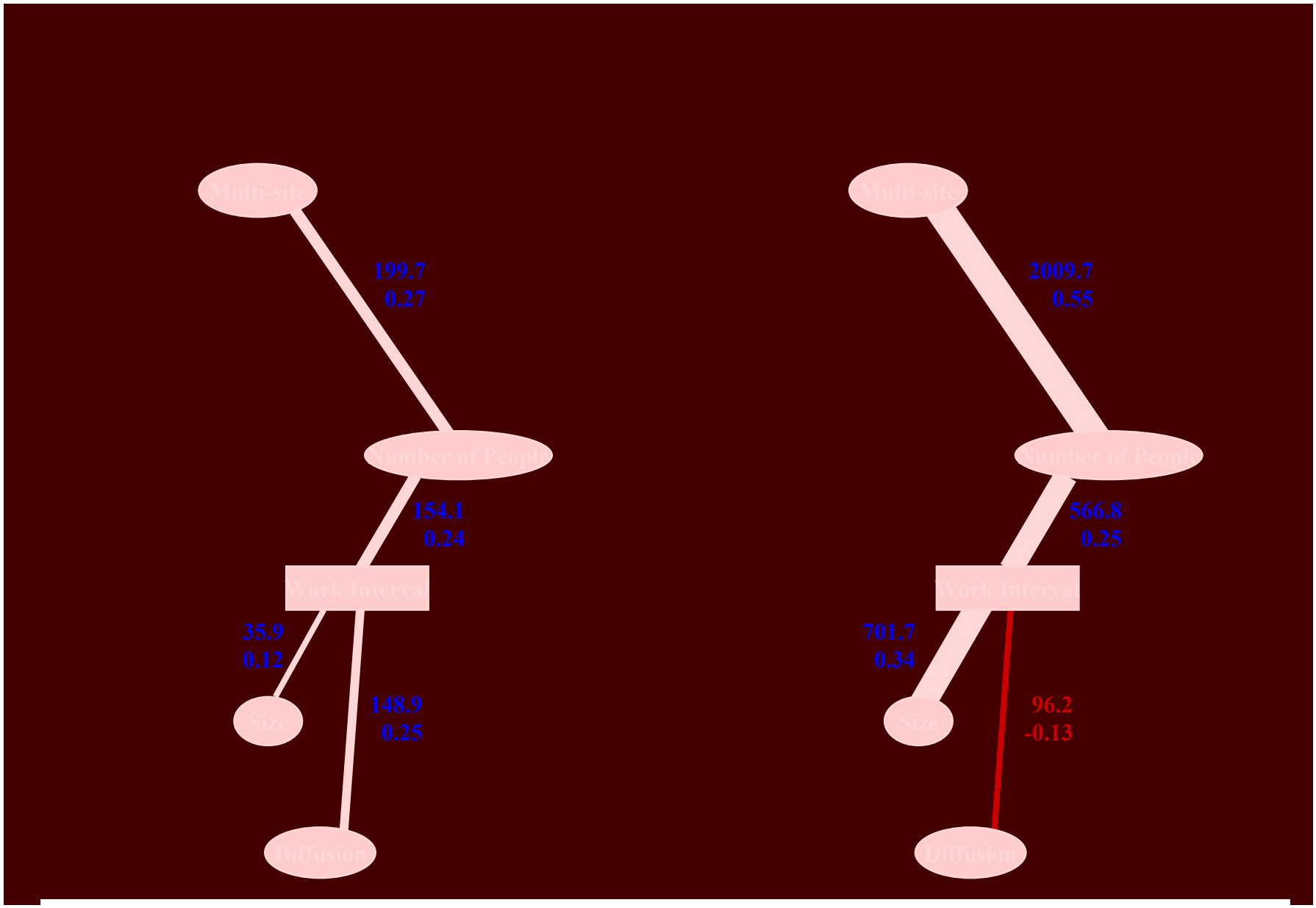
H4 Multi-site MRs involve more people, take longer

Graphical model of work interval for Network Element A





Graphical model of work interval for Network Element A (left) and B (right)



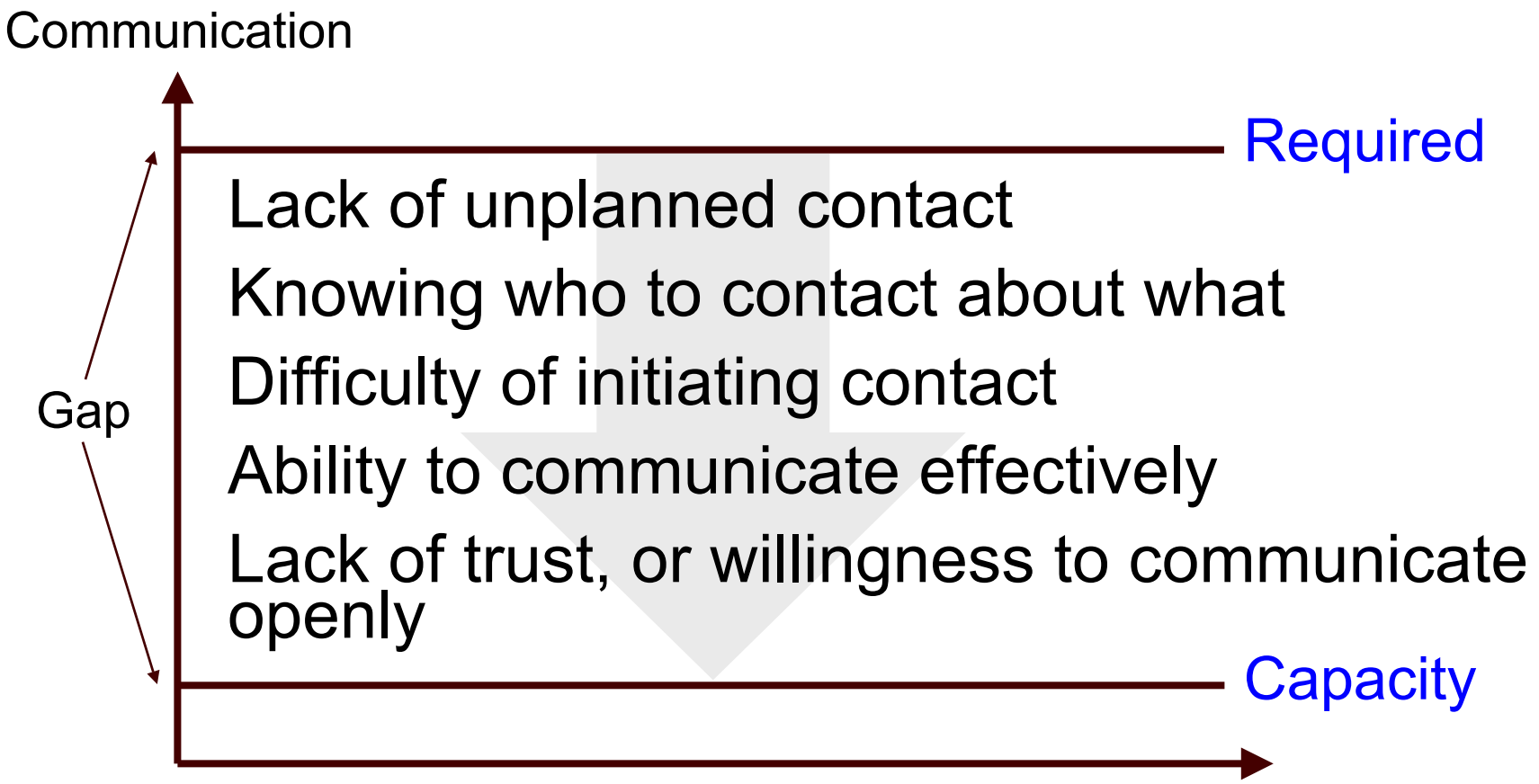
Graphical model of work interval for Network Element A (left) and B (right)

MR Interval
Distance Requires More People?

- MR is assigned to “owner” who recruits others
- Multi-site requires more people?
  - Who is the right expert?
  - Can MR owner get “right person” to do the work?
  - “Partial” expertise, several people
  - Correct errors

Co-located Development

Multi-site Development



Effects

- Cannot deal effectively with the unexpected, re-negotiating commitments
- Issue resolution paralysis
- Very difficult to stay “in the loop”
- Ineffective collaborative sessions

Solutions

Tools

TeamPortal

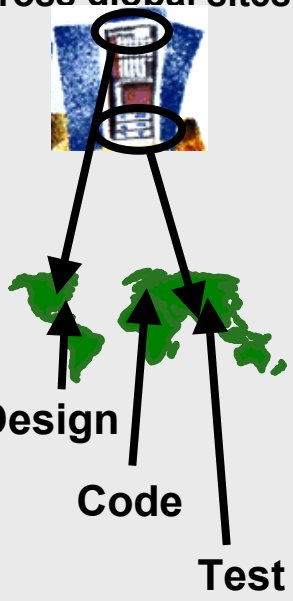
Rear View Mirror

CalendarBot

Experience Browser

Models of Development

How to distribute work across global sites.



Best Practices

Planning Travel

xxxxxxxxxxxxxxxx
xxxxxx

Establishing Liaisons

xxxxxx xxxxx xxxxxxxxxxxxxx
xxxxx xxxx
xxxxxxxxxxxxxxx

Building Trust

xxxxxx xxxxx
xxxxxxxxx

Communication Etiquette

xxxxxxxxxxxxxxxxxx
xxxxxxx xxxxxx

Preventing Delay

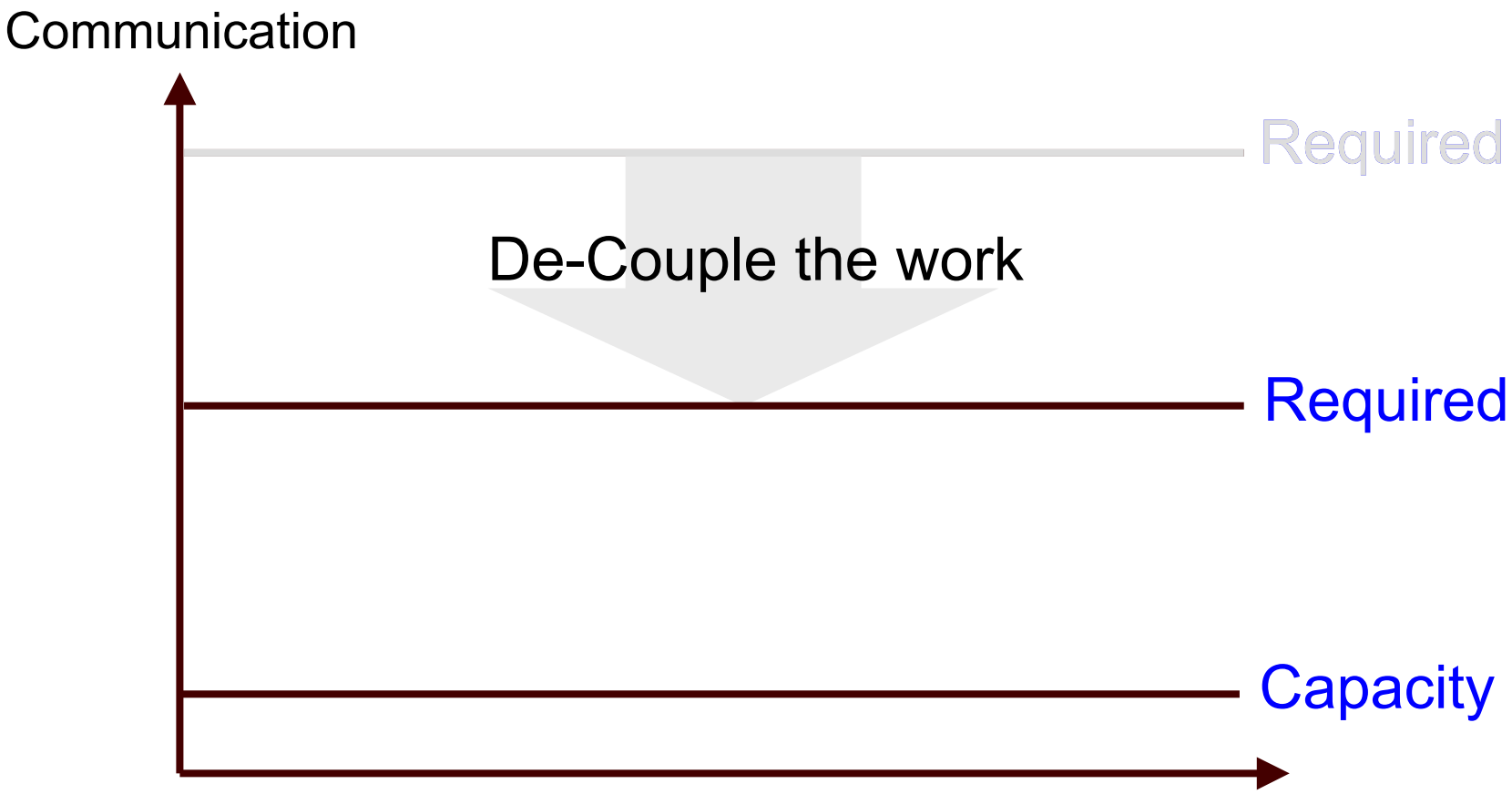
xxxxxxx xxx xxxxx
xxxxxxxxxx
xxxxx xxxxxxx

Using Commercial Tools

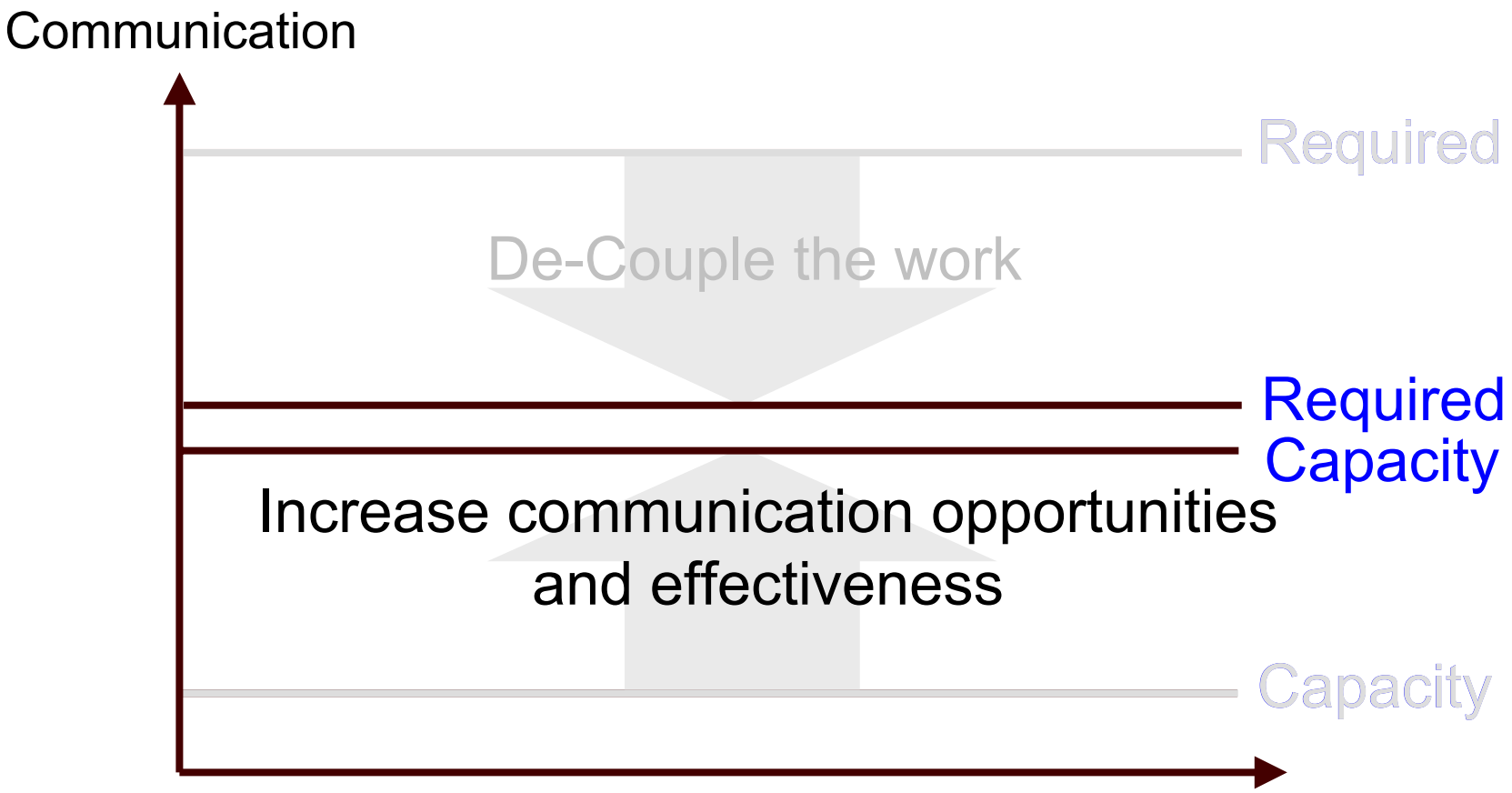
xxxxxxxx
xxxxxx xxxx xxxxxx
xxxxx xxxxx

Global Development Solutions

Bridging the Gap - 1



Bridging the Gap - 2



GSD -- Laboratory for Studying Coordination in Software Engineering

- Effects of missing coordination processes are highly visible, e.g.,
  - Informal communication
  - Practices surrounding process
  - Weaknesses, failures in coordination mechanisms
- Much related theory
  - Coordination theory (e.g., Malone)
  - Distributed cognition (e.g., Hutchins)
  - Multi-agent planning (e.g., Durfee)
  - Modularity (e.g., Parnas)
  - Organizational structure (e.g., Conway)

Coordination in Software Development: Intuitions

- Software projects progress by making decisions
- Decisions constrain other decisions
- Difficult to bring all constraints to bear on decision-making
- Difficult to know current state of decision-making

Toward a Software Engineering Theory of Coordination

- Initial goal
  - Rigorous formulation of Parnas and Conway principles
  - Preliminary empirical test
  - Define a research program

Definitions

- Software project is a set of decisions X_i
- Design space X_{ij}
- Goal space -- subset of design space that leads to satisfaction of requirements
- Constraints over decisions -- implicitly defined by feasibility function:

$f(x_{1,k_1},...,x_{n,k_n})$ ={1 iff product satisfies requirements,
0 otherwise}

Effects of a Decision

Feasible choices (FC) for decision X:

$$x_{kj} : \forall i \exists j(i) \text{ such that } f(x_{1\,j(1)}, \ldots, x_{kj(k)}, \ldots, x_{nj(n)}) = 1$$

Effects of a decision $k : X_k = x_k$ on a decision l

$$E(X_l \mid X_k = x_k) =$$

$$FC(X_l) - FC(X_l \mid X_k = x_k)$$

"Laws" of Coordination

Parnas, information hiding:

$$\forall i, j, k : X_i \in M_j, X_k \notin M_j, E(X_i | X_k) = \emptyset$$

Where M_j is a clump of decisions defined by module boundaries.

Conway's Law:

$$\forall j \exists i : M_j \subset T_i,$$

Where T_i is a clump of decisions induced by organizational structure.

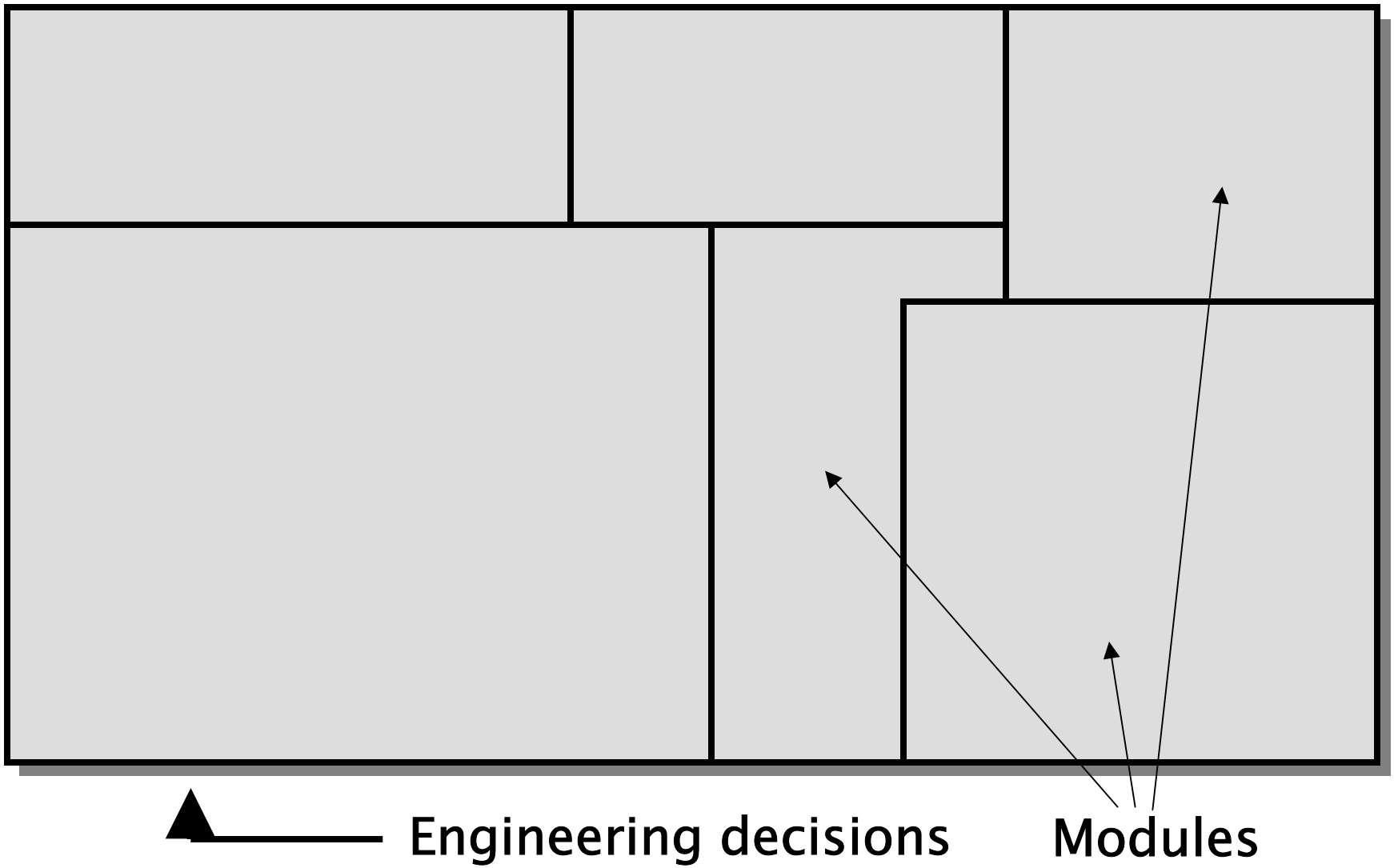
Assumptions to Generate Predictions

- Effects of infeasible choices
  - defects, faults, errors, failure to complete project
  - rework
    - longer cycle time
    - lower productivity
- Feasible choices more likely when
  - the decisions are made by one person
  - the decisions are made close together in time
  - decisions are made by people in frequent communication with each other
  - the set of choices already made on a project is highly visible to decision makers
  - the constraints that bear on a decision are highly visible to decision makers

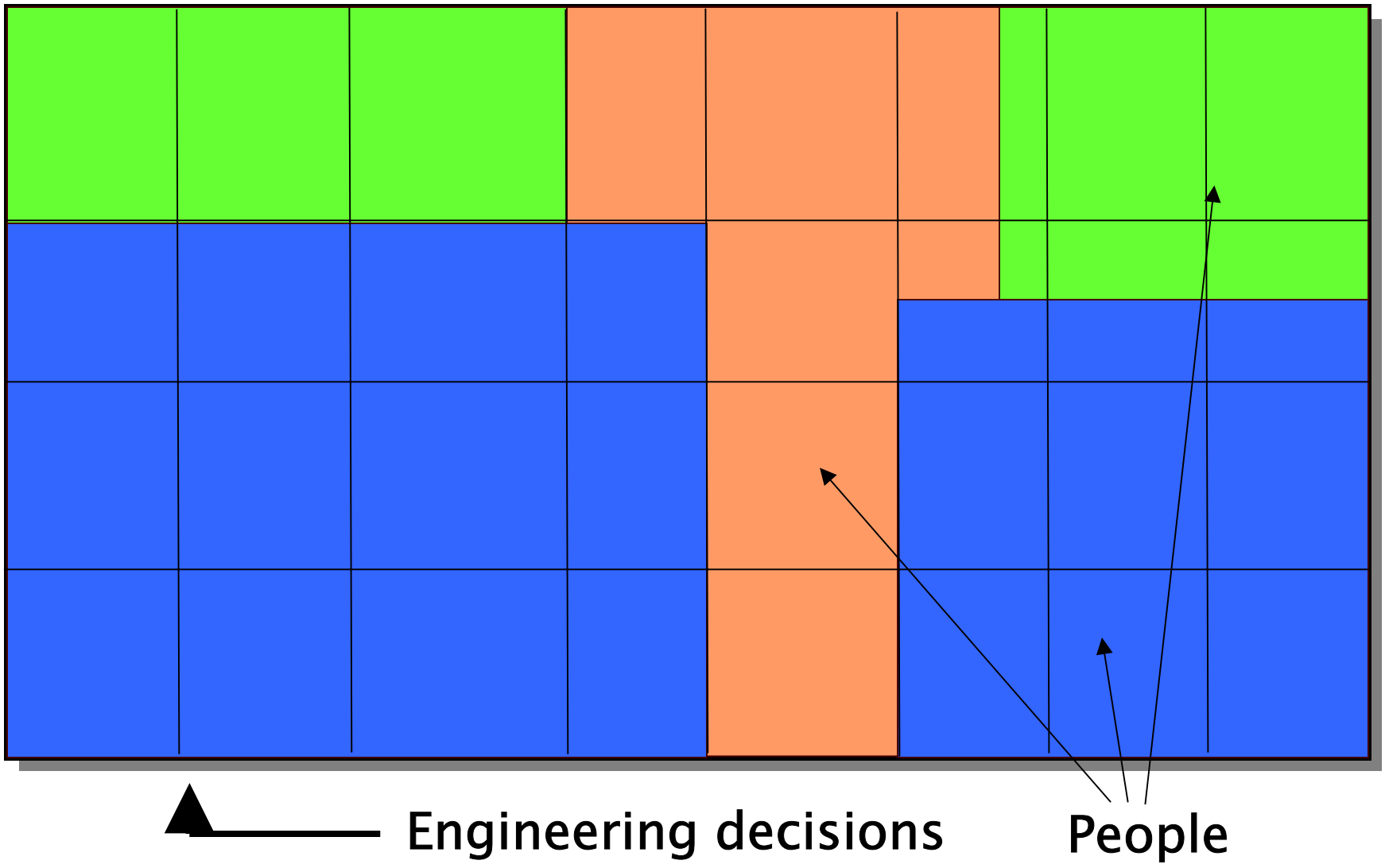
Hypothesis Testing Strategy

- Partition decisions three different ways:
  - Product structure (e.g., modules)
  - Organizational structure (e.g., teams, individuals)
  - Work units (e.g., modification requests)

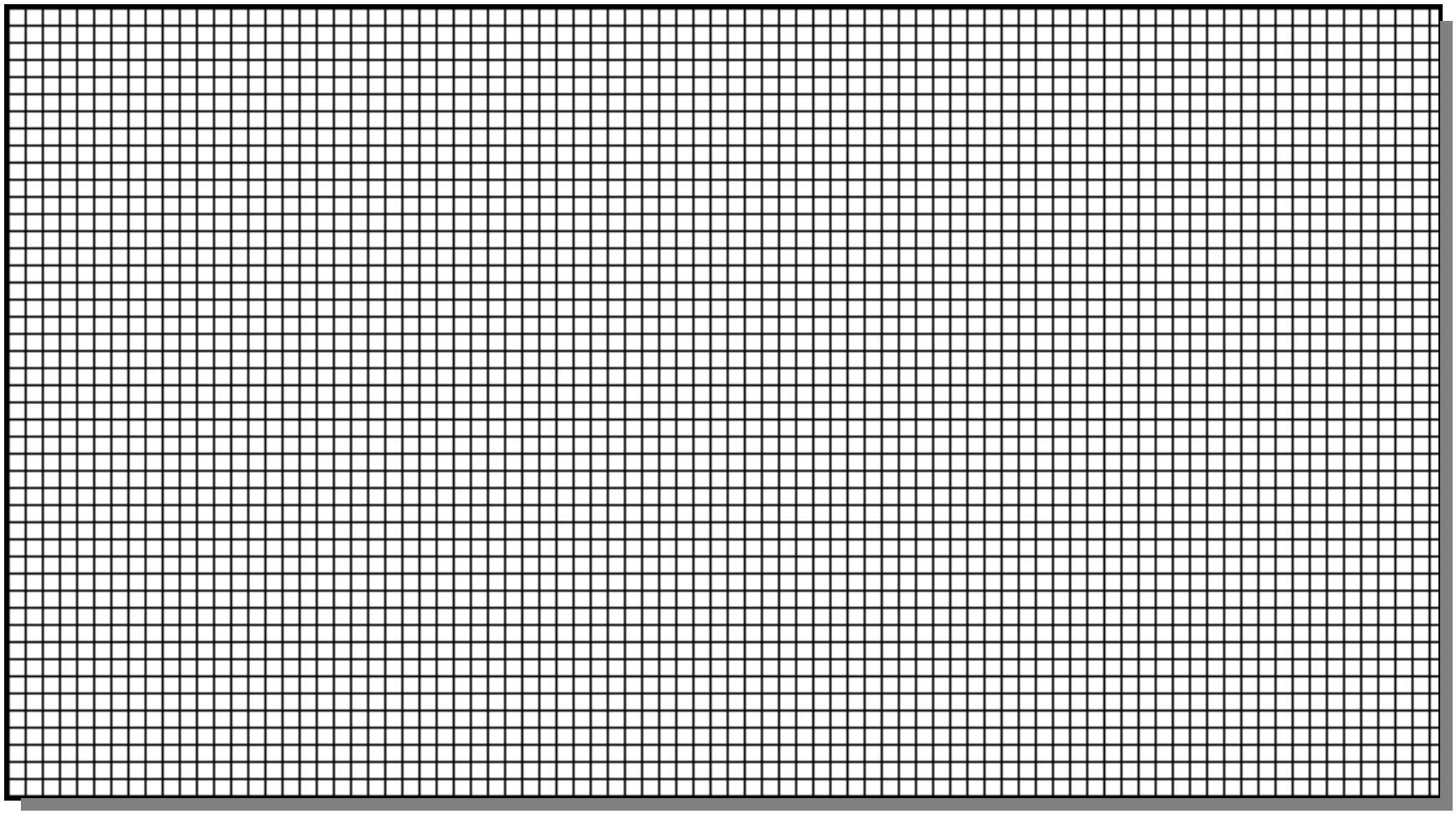
Product Structure Partitions



Organizational Structure Partitions

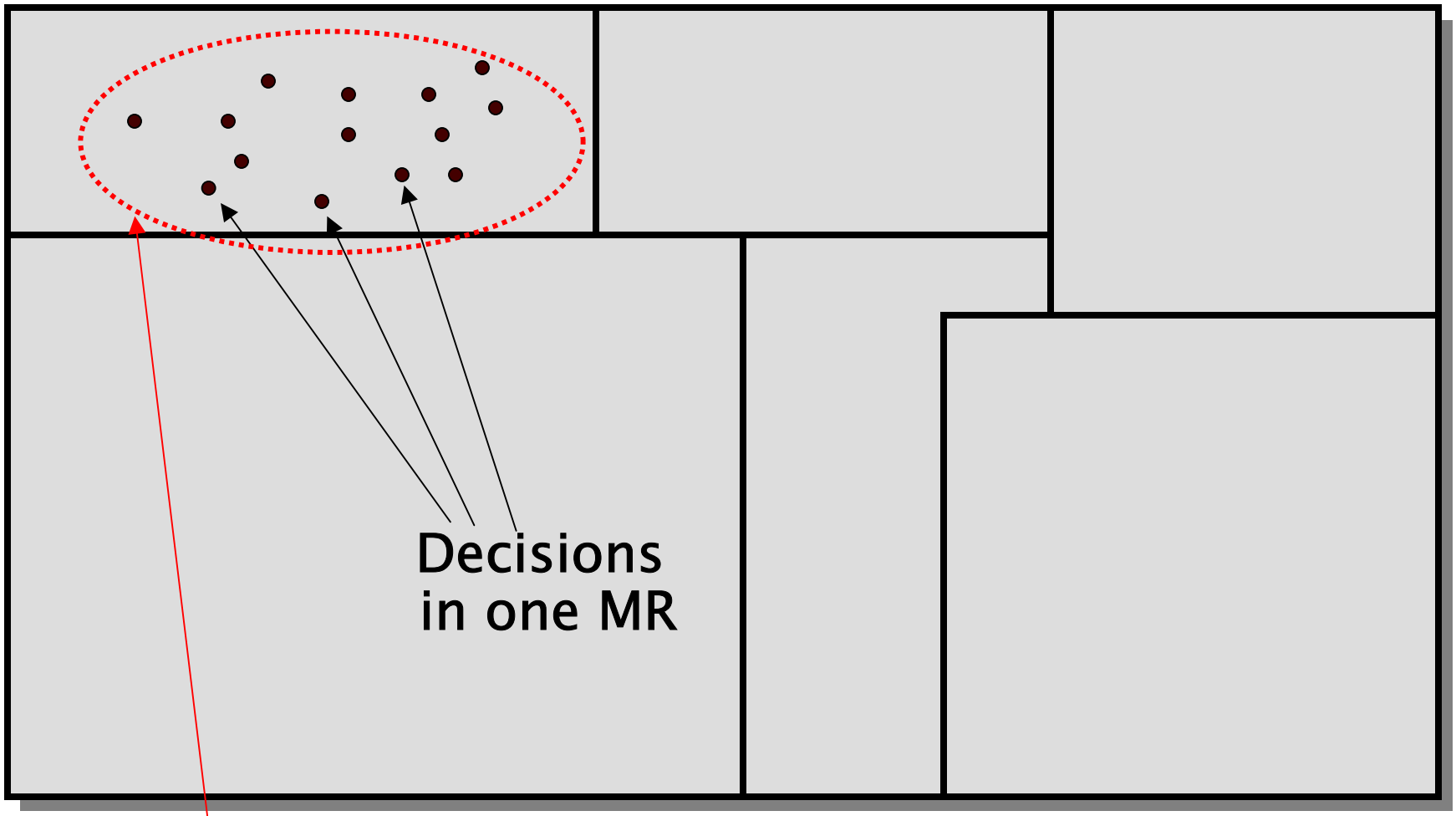


Work Item Partition



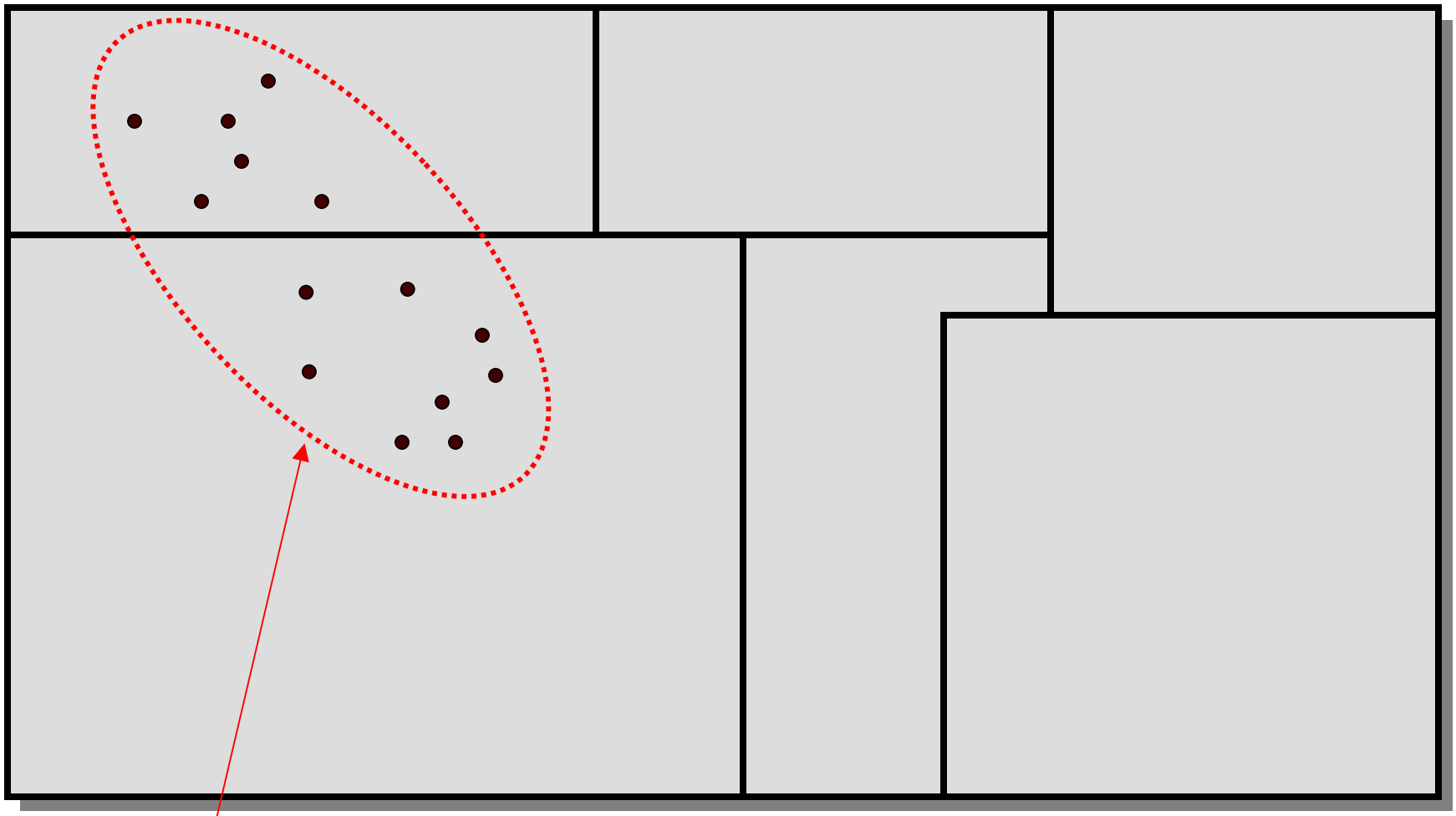
Engineering decisions

Hypothesis Testing Strategy



MR: Set of mutually-constraining decisions

Hypothesis Testing Strategy



MR: Set of mutually-constraining decisions

Product Structure (Parnas) Hypothesis H1

- Empirical workflow graph
  - Partitioned code into two empirically-derived modules
  - Nodes are files
  - Edges connect files modified in same MR
- *H2: Modification requests that require work in different modules will have longer cycle times than modification requests that require work in only a single module.*
- Hypothesis supported, controlling for other relevant variables in a regression model

Organizational Structure (Conway) Hypothesis H2

- Empirical workflow graph
  - Nodes are people
  - Arcs represent sequence of work within a work item (modification request)
- *H1: Developers with more people assigning work to them will have lower productivity.*
- Hypothesis supported, controlling for other relevant variables in a regression model

Implications for GSD

- Focus on dependencies and coordination
- Crucial role of informal communication
- Visibility of decisions
- Role of product lines, architectures, design methods