

1

2

3

4

Philipps



Universität

Marburg

Variability Mining



Christian Kästner
Philipps University Marburg

Joint work with Paolo Giarusso, Tillmann Rendel, Sebastian Erdweg, Klaus Ostermann, Steffen Haase, Alexander Dreiling, Sven Apel, Olaf Lessenich, Slawomir Duszynski

From Legacy Code

to Disciplined Product-Line Implementations

Extraction

Transformation

Implementation

Domain knowledge

Legacy System with #ifdef Variability

static variability extr.

Legacy System with Runtime Parameters

dynamic variability extr.

Legacy System without Variability

Feature Mining

Family of Similar Legacy Systems

Legacy 1

Legacy 2a

Legacy 2b

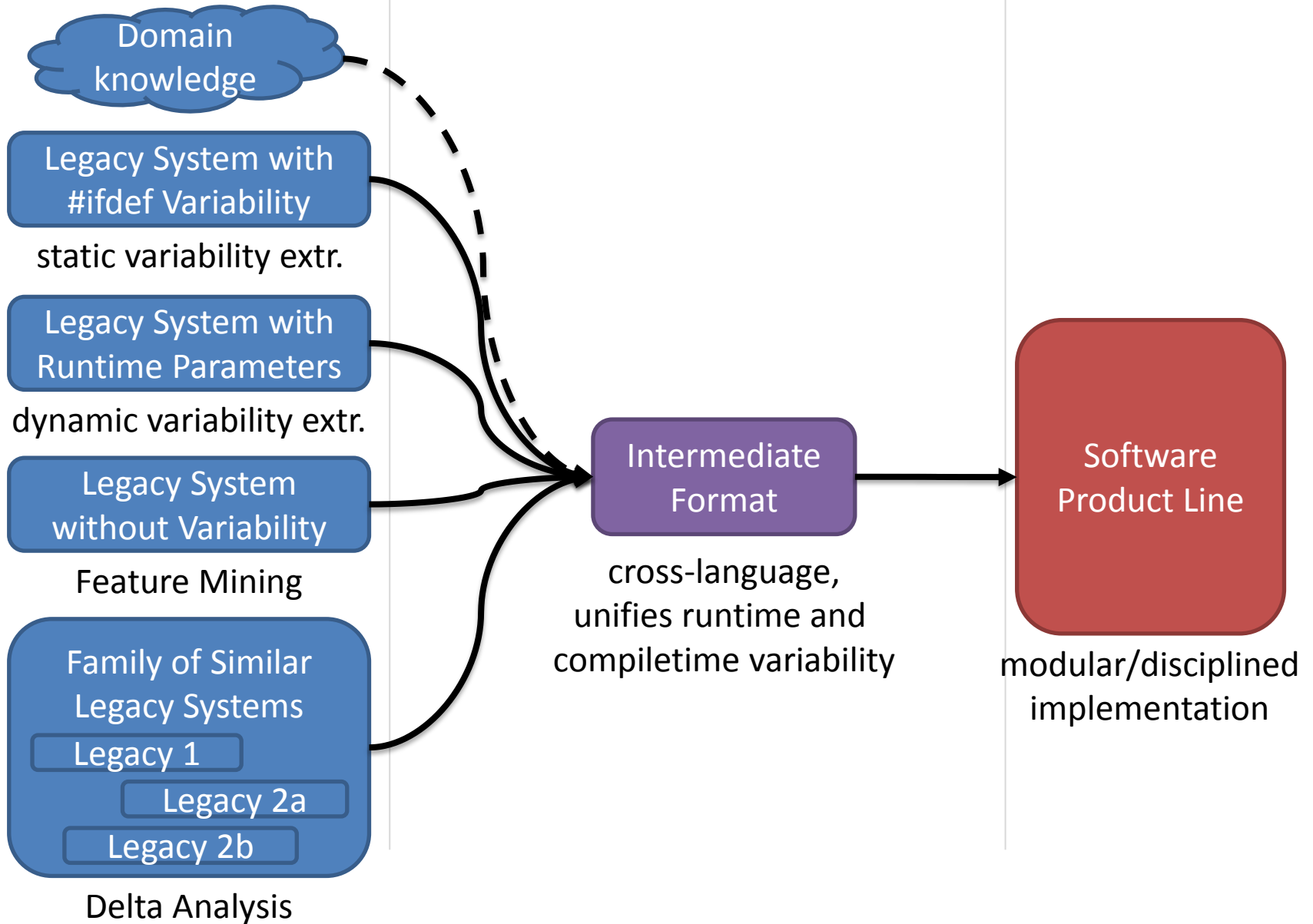
Delta Analysis

Intermediate Format

cross-language,
unifies runtime and
compiletime variability

Software Product Line

modular/disciplined
implementation



Extraction

Transformation

Implementation

Domain
knowledge

Legacy System with
#ifdef Variability

static variability extr.

Legacy System with
Runtime Parameters

dynamic variability extr.

Legacy System
without Variability

Feature Mining

Family of Similar
Legacy Systems

Legacy 1

Legacy 2a

Legacy 2b

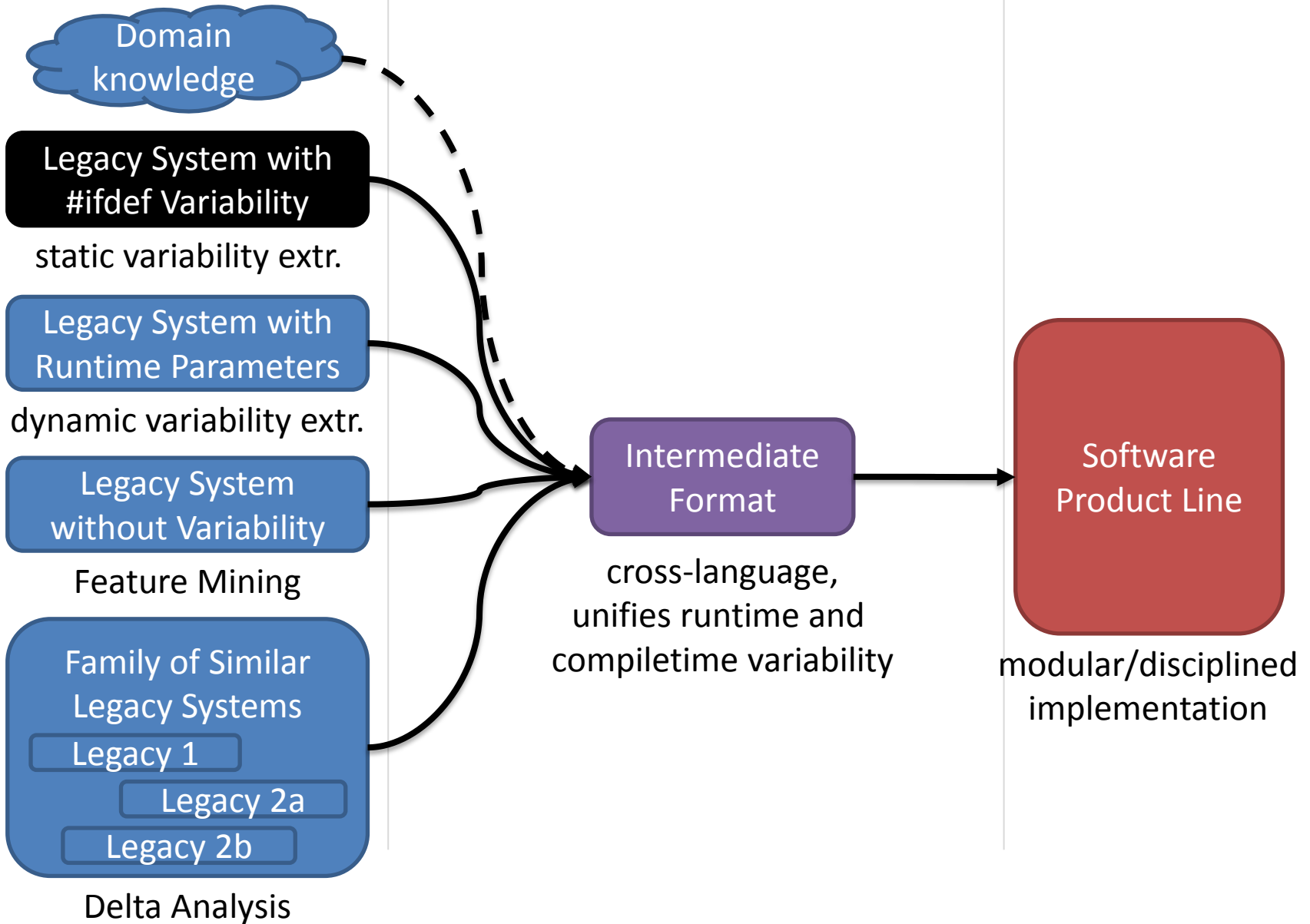
Delta Analysis

Intermediate
Format

cross-language,
unifies runtime and
compiletime variability

Software
Product Line

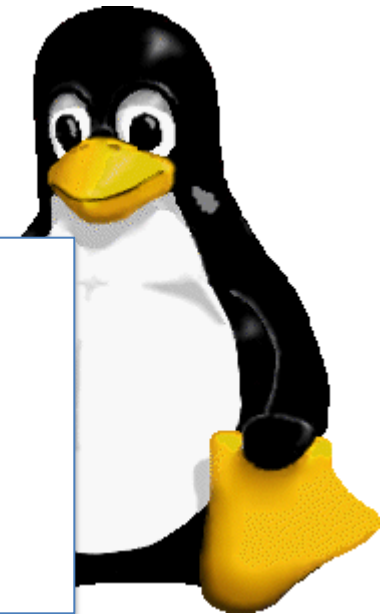
modular/disciplined
implementation



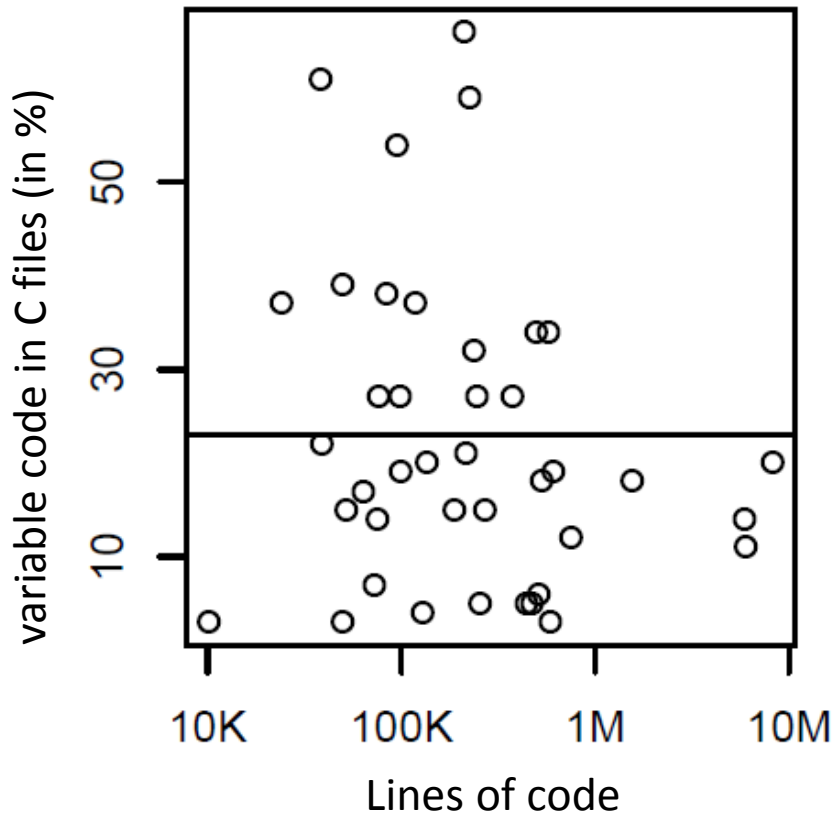
Lexical Preprocessors

10,000 features, 6 million lines of C code

```
static void rt_mutex_init_task(struct task_struct *p)
{
    raw_spin_lock_init(&p->pi_lock);
#ifdef CONFIG_RT_MUTEXES
    plist_head_init_raw(&p->pi_waiters, &p->pi_lock);
    p->pi_blocked_on = NULL;
#endif
}
```



40 Open-Source C Projects



apache, berkely db,
 cherokee, clamav, dia,
 emacs, freebsd, gcc,
 ghostscript, gimp, glibc,
 gnumeric, gnuplot, irssi,
 libxml, lighttpd, linux, lynx,
 minix, mplayer, mpsolve,
 openldap, opensolaris,
 openvpn, parrot, php, pidgin,
 postgresql, privoxy, python,
 sendmail, sqlite, subversion,
 sylpheed, tcl, vim, xfig,
 xine-lib, xorg-server, xterm


```
    int n = NUM2INT(num);  
#ifndef FEAT_WINDOWS  
    w = curwin;  
#else  
    for (w = firstwin; w != NULL;  
         w = w->w_next, --n)  
#endif  
    if (n == 0)  
        return window_new(w);
```



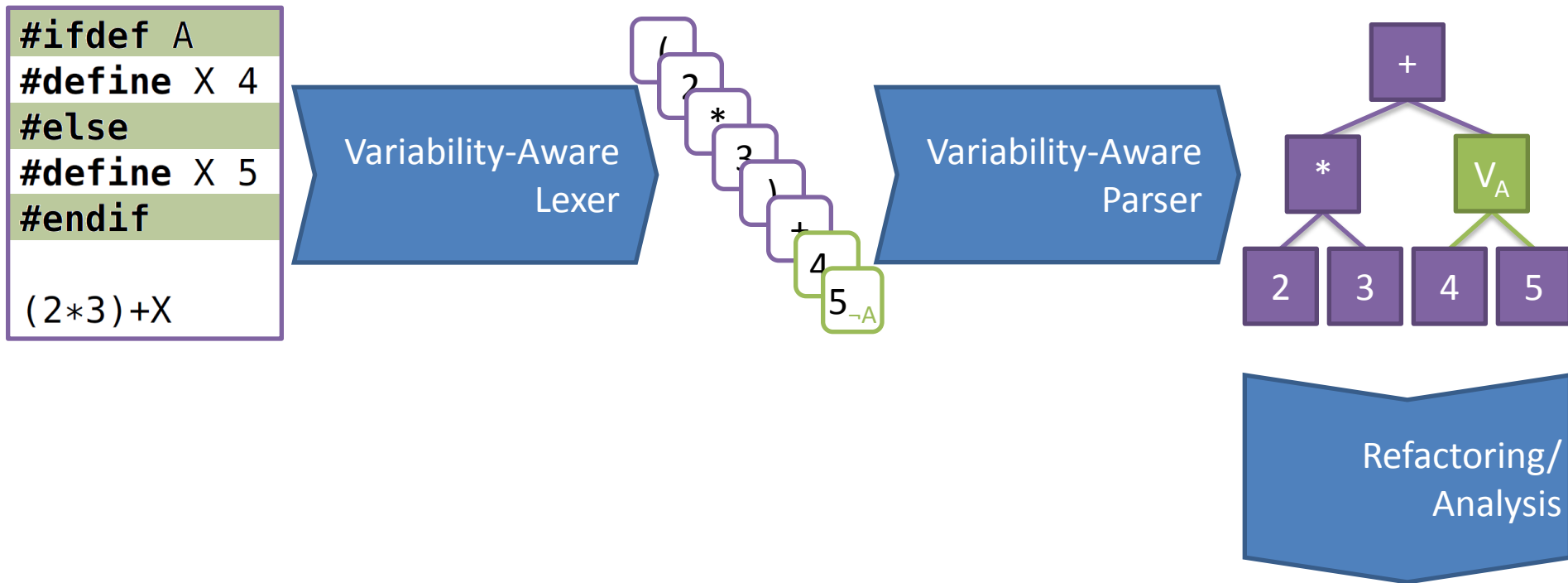
```
int n = M
#ifndef F
w = curw
#else
for (w =
w = w
#endif
if (n =
return
```

```
int put_eol(fd)
FILE *fd;
{
    if (
#ifdef USE_CRNL
        (
#ifdef MKSESSION_NL
            !mksession_nl &&
        )
#endif
        (putc('\r', fc) < 0) ||
    )
#endif
        (putc('\n', fd) < 0)
        return FAIL;
    return OK;
}
```



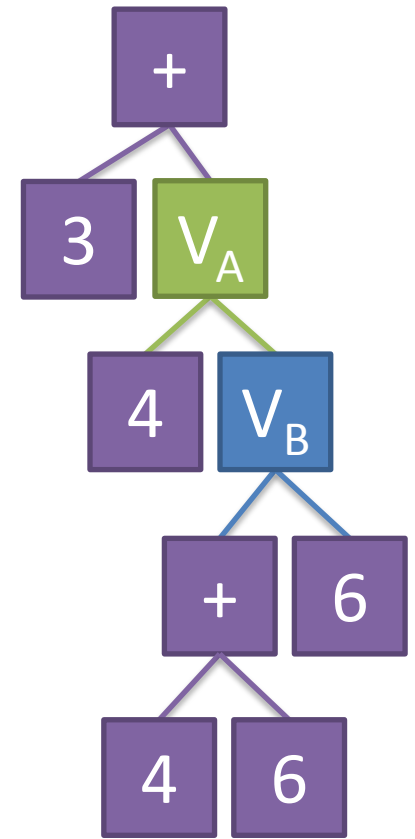
TypeChef

[OOPSLA'11]



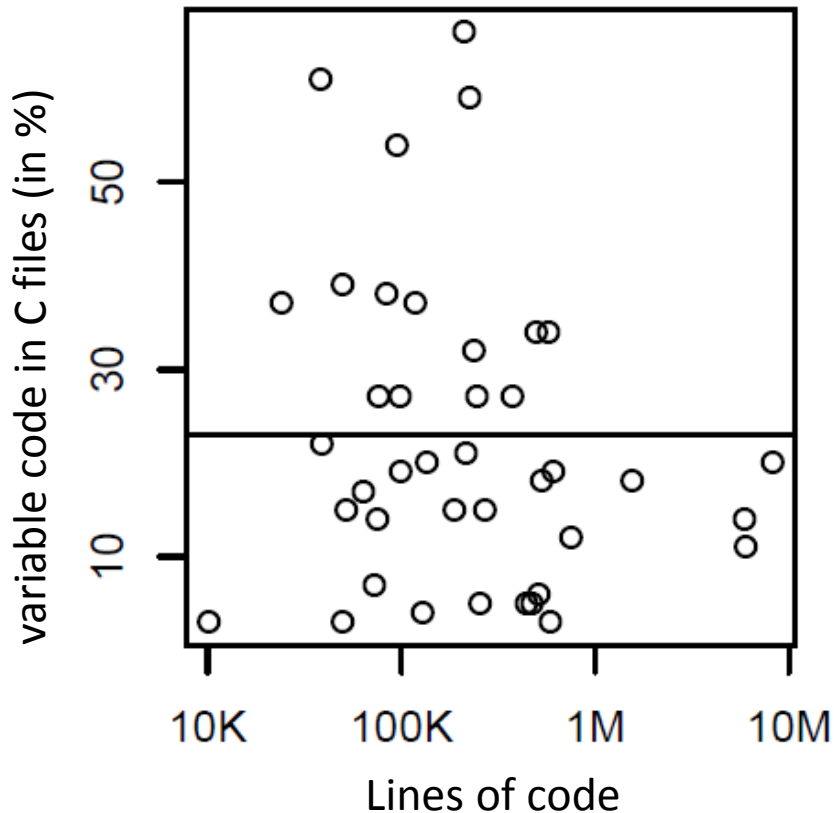


```
(3+
#ifdef A
4
#else
(
#ifdef B
4+
#endif
6)
#endif
)
```



Variability-Aware Parsing

40 Open-Source C Projects



apache, berkely db,
 cherokee, clamav, dia,
 emacs, freebsd, gcc,
 ghostscript, gimp, glibc,
 gnumeric, gnuplot, irssi,
 libxml, lighttpd, linux, lynx,
 minix, mplayer, mpsolve,
 openldap, opensolaris,
 openvpn, parrot, php, pidgin,
 postgresql, privoxy, python,
 sendmail, sqlite, subversion,
 sylpheed, tcl, vim, xfig,
 xine-lib, xorg-server, xterm

Extraction

Transformation

Implementation

Domain
knowledge

Legacy System with
#ifdef Variability
static variability extr.

Legacy System with
Runtime Parameters
dynamic variability extr.

Legacy System
without Variability
Feature Mining

Family of Similar
Legacy Systems

Legacy 1

Legacy 2a

Legacy 2b

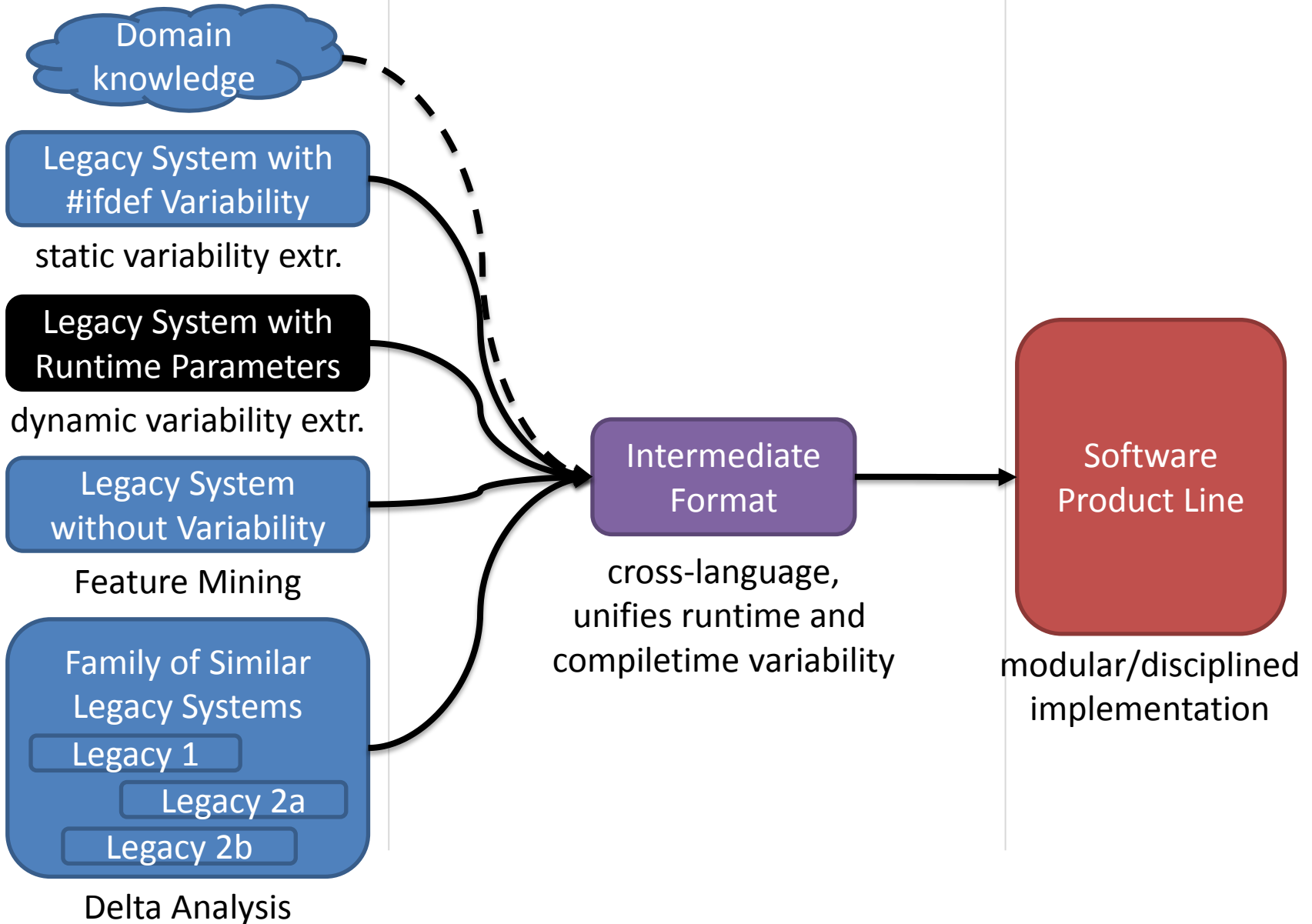
Delta Analysis

Intermediate
Format

cross-language,
unifies runtime and
compiletime variability

Software
Product Line

modular/disciplined
implementation



Administrator: C:\Windows\system32\cmd.exe

```
C:\Users\ckaestne>grep --help
```

```
Aufruf: grep [OPTION]... MUSTER [DATEI]...
```

```
Search for PATTERN in each FILE or standard input.
```

```
PATTERN is, by default, a basic regular expression (BRE).
```

```
Example: grep -i 'hello world' menu.h main.c
```

```
Regex selection and interpretation:
```

-E, --extended-regexp	PATTERN is an extended regular expression (ERE)
-F, --fixed-strings	PATTERN is a set of newline-separated fixed strings
-G, --basic-regexp	PATTERN is a basic regular expression (BRE)
-P, --perl-regexp	PATTERN is a Perl regular expression
-e, --regexp=PATTERN	use PATTERN for matching
-f, --file=FILE	obtain PATTERN from FILE
-i, --ignore-case	ignore case distinctions
-w, --word-regexp	force PATTERN to match only whole words
-x, --line-regexp	force PATTERN to match only whole lines
-z, --null-data	a data line ends in 0 byte, not newline

```
Verschiedenes:
```

-s, --no-messages	Fehlermeldungen unterdrücken.
-v, --invert-match	Nicht-passende Zeilen anzeigen.
-V, --version	Versionsnummer ausgeben und beenden.
--help	Diese Hilfe ausgeben und beenden.
--mmap	Wenn möglich, Eingabe in den Speicher mappen.

Forward Slicing

Interprocedural static analysis classifies into

- (a) belongs to feature
- (b) does not belong to f.
- (c) unsure

```
prepend_default_options (getenv ("GREP_OPTIONS"), &argc, &argv);  
[...]  
while ((opt = get_nondigit_option (argc, argv, &default_context)) != -1)  
    switch (opt)  
    [...]  
    case 'i':  
    case 'y': /* For old-timers . . . */  
        match_icode = 1;  
        break;  
[...]
```

```
if (match_icode )  
    {  
        int i = lim - beg;  
  
        ibeg = buf = xmalloc(i);  
        while (--i >= 0)  
            buf[i] = tolower((unsigned char) beg[i]);  
    }  
else  
    {  
        buf = NULL;  
        ibeg = beg;  
    }
```


Source Code Viewer

Features File Structure

Example

- src
 - dfa.c
 - dfasearch.c
 - dosbuf.c
 - egrep.c
 - fgrep.c
 - grep.c
 - kwsearch.c
 - kwset.c
 - main.c
 - pcrsearch.c
 - searchutils.c

main.c

```
825 }
826
827 static void
828 prline (char const *beg, char const *lim, int sep)
829 {
830     int matching;
831     const char *line_color;
832     const char *match_color;
833
834     if (!only_matching)
835         print_line_head(beg, lim, sep);
836
837     matching = (sep == SEP_CHAR_SELECTED) ^ !!out_invert;
838
839     if (color_option)
840     {
841         line_color = ( (sep == SEP_CHAR_SELECTED)
842                     ^ (out_invert && (color_option < 0)))
843                     ? selected_line_color : context_line_color;
844         match_color = (sep == SEP_CHAR_SELECTED)
845                     ? selected_match_color : context_match_color;
846     }
847     else
848         line_color = match_color = NULL; /* Shouldn't be used. */
849
850     if ( (only_matching && matching)
851         || (color_option && (*line_color || *match_color)))
852     {
853         /* We already know that non-matching lines have no match (to colorize). */
854         if (matching && (only_matching || *match_color))
855             beg = print_line_middle(beg, lim, line_color, match_color);
856
857         /* FIXME: this test may be removable. */
858         if (!only_matching && *line_color)
859             beg = print_line_tail(beg, lim, line_color);
860     }
861
862     if (!only_matching && lim > beg)
863         fwrite (beg, 1, lim - beg, stdout);
```

Feature Model

- CONFIG_match_words
- CONFIG_IFPOSSIBLE_match_words
- CONFIG_IFEIPossible_match_wo...
- CONFIG_match_lines
- CONFIG_IFPOSSIBLE_match_lines
- CONFIG_IFEIPossible_match_lines
- CONFIG_IFPOSSIBLE_match_icode
- CONFIG_match_icode
- CONFIG_IFEIPossible_match_icode

Colors

ColorAssignments

Uses interprocedural static analysis

```
prepend_default_options (getenv ("GREP_OPTIONS"), &argc, &argv);  
[...]  
while ((opt = get_nondigit_option (argc, argv, &default_context)) != -1)  
    switch (opt)  
    [...]  
    case 'i':  
    case 'y': /* For old-timers . . . */  
        match_licase = 1;  
        break;
```

[...]

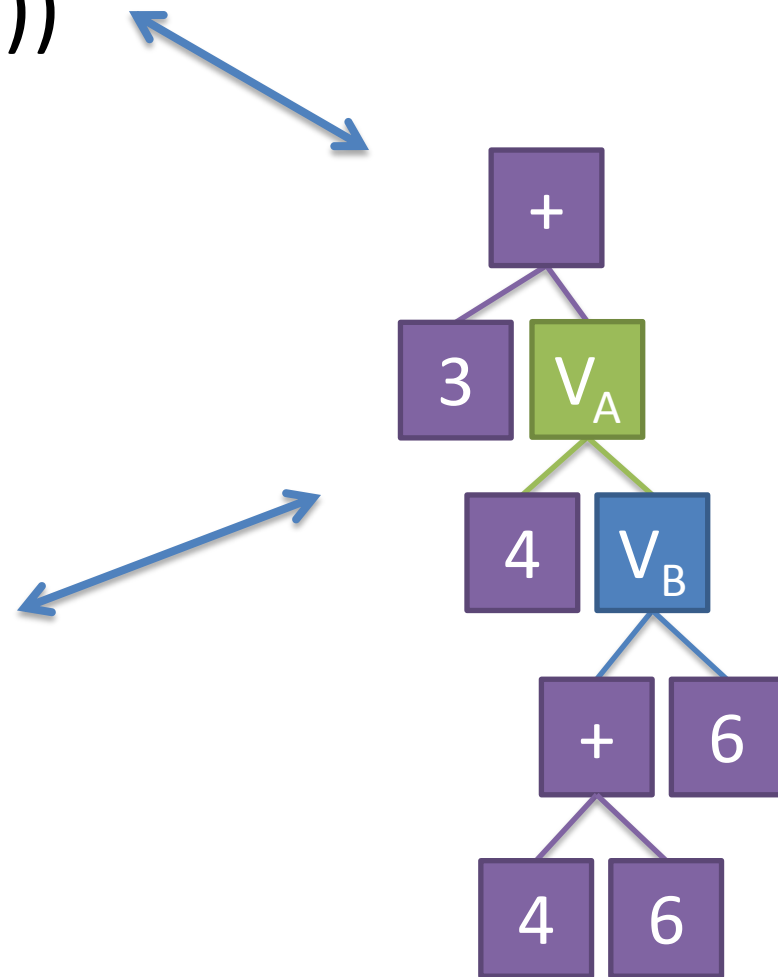
```
if (match_licase )  
    {  
        int i = lim - beg;  
  
        ibeg = buf = xmalloc(i);  
        while (--i >= 0)  
            buf[i] = tolower((unsigned char)  
    }  
else  
    {  
        buf = NULL;  
        ibeg = beg;  
    }
```

```
#IFDEF FEAT_ICASE  
    if(MB_CUR_MAX == 1) {  
        int i = lim - beg;  
  
        ibeg = buf = xmalloc(i);  
        while (--i >= 0)  
            buf[i] = tolower((unsigned char)  
    } else {  
        buf = NULL;  
        ibeg = beg;  
    }  
#ELSE  
    buf = NULL;  
    ibeg = beg;  
#ENDIF
```

bool A, B;

3 + (A ? 4 : (B ? 4+6 : 6))

```
(3+  
#ifdef A  
4  
#else  
(  
#ifdef B  
4+  
#endif  
6)  
#endif  
)
```



	grep	vsftpd	nglRCd
Version	2.6.1	2.3.4	0.17.1
Code Files in whole project	238	97	84
LOC of whole project	≈ 39000	≈ 15500	≈ 16500
Files in main source folder	19	97	65
LOC of main source folder	≈ 6100	≈ 15500	≈ 14000
# of Functions in analysis	134	738	537

Extraction

Transformation

Implementation

Domain knowledge

Legacy System with #ifdef Variability

static variability extr.

Legacy System with Runtime Parameters

dynamic variability extr.

Legacy System without Variability

Feature Mining

Family of Similar Legacy Systems

Legacy 1

Legacy 2a

Legacy 2b

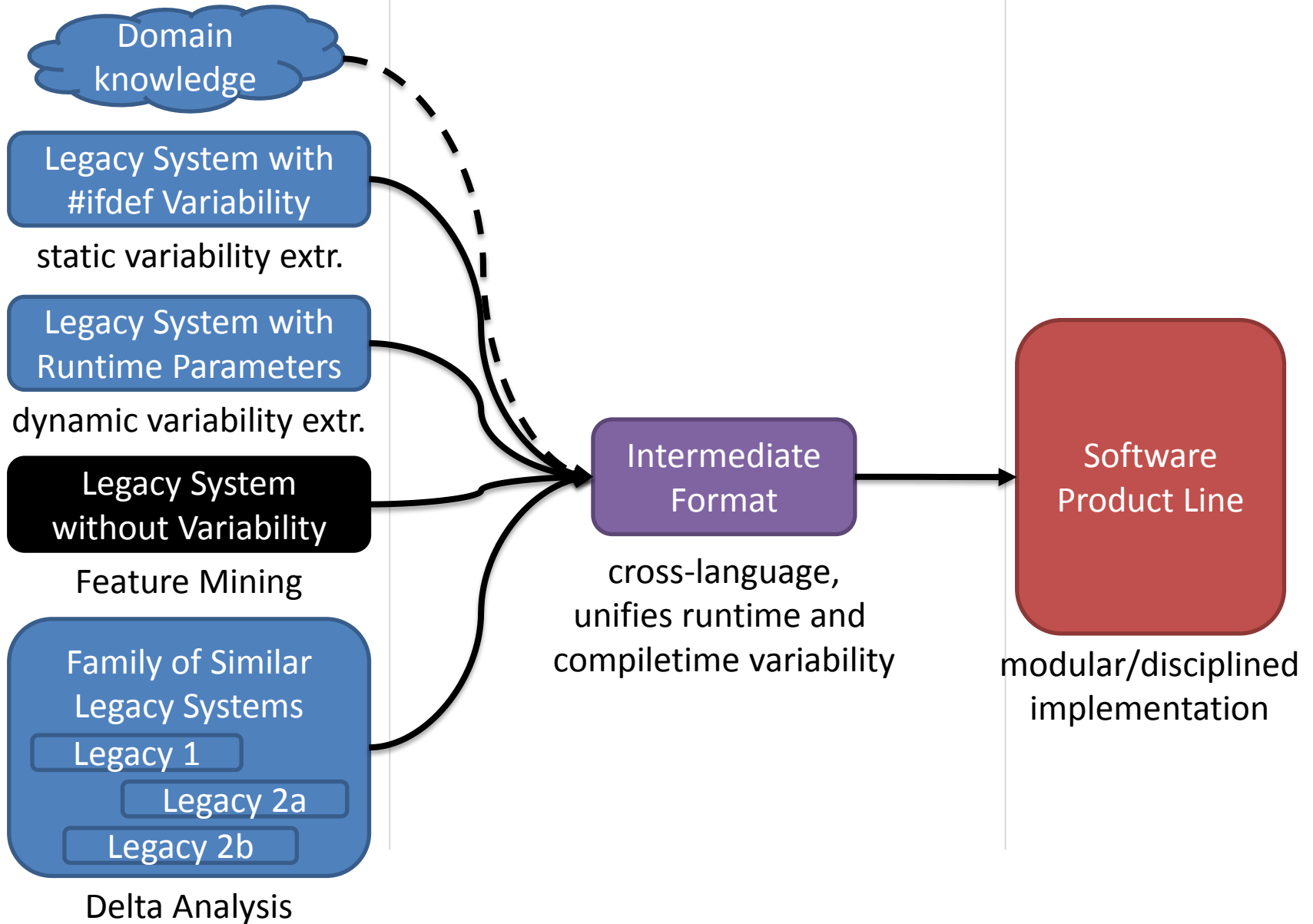
Delta Analysis

Intermediate Format

cross-language,
unifies runtime and
compiletime variability

Software Product Line

modular/disciplined
implementation



```
class Stack {
    int size = 0;
    Object[] elementData = new Object[maxSize];
    boolean transactionsEnabled = true;

    void push(Object o) {
        Lock l = lock();
        elementData[size++] = o;
        unlock(l);
    }
    Object pop() {
        Lock l = lock();
        Object r = elementData[--size];
        unlock(l);
        return r;
    }
    Lock lock() {
        if (!transactionsEnabled) return null;
        return Lock.acquire();
    }
    void unlock(Lock lock) { /*...*/ }
    String getLockVersion() { return "1.0"; }
}
class Lock { /*...*/ }
```

```
class Stack {
    int size = 0;
    Object[] elementData = new Object[maxSize];
    boolean transactionsEnabled = true;

    void push(Object o) {
        Lock l = lock();
        elementData[size++] = o;
        unlock(l);
    }
    Object pop() {
        Lock l = lock();
        Object r = elementData[--size];
        unlock(l);
        return r;
    }
    Lock lock() {
        if (!transactionsEnabled) return null;
        return Lock.acquire();
    }
    void unlock(Lock lock) { /*...*/ }
    String getLockVersion() { return "1.0"; }
}
class Lock { /*...*/ }
```



seed

based on

Impact Analysis /
Concern Location

```
class Stack {
    int size = 0;
    Object[] elementData = new Object[maxSize];
    boolean transactionsEnabled = true;

    void push(Object o) {
        Lock l = lock();
        elementData[size++] = o;
        unlock(l);
    }
    Object pop() {
        Lock l = lock();
        Object r = elementData[--size];
        unlock(l);
        return r;
    }
    Lock lock() {
        if (!transactionsEnabled) return null;
        return Lock.acquire();
    }
    void unlock(Lock lock) { /*...*/ }
    String getLockVersion() { return "1.0"; }
}
class Lock { /*...*/ }
```

Consistency Criteria
Granularity

seed

Type System
calls, references

Topology Analysis
call hierarchy

Information Retrieval
name similarity

based on
Impact Analysis /
Concern Location

LEADT (Location, Expansion, And Documentation Tool)

Resource - MobileMedia7-CIDE-Compliant/src/ubc/midp/mobilephoto/core/ui/controller/PhotoViewController.java - Eclipse Platform

File Edit Navigate Search Project JReps Colors Run Window Help

Project Explorer

- src
 - lanca.midp.mobilephoto
 - ubc.midp.mobilephoto
 - core
 - comms
 - threads
 - ui
 - util
 - sms
 - JRE System Library [JavaSE-1
 - Mobile
 - model.colors
 - model.m

```

deCopyPhoto
d in scenario 06 */
Data().addMediaData(imageData, albumname);

ImageDataException e) {
= null;
eof ImagePathNotValidException)
new Alert("Error", "The path is not valid", nul
new Alert("Error", "The image file format is no
isplay(midlet).setCurrent(alert, Display.getDis
Timeout(5000);

```

Name

- SimNam: albumname
- X: Direct Relations
 - belongs to
 - VaDeFa: albumname
 - X: Transpose Relations
 - *requires
 - local variable access declared
 - local variable accessed by
 - local variable transitively acce
 - parameter accessed by
 - MetInv: addMediaData

Clear

FeatureManager

NonFeatureManager

FeatureManager

base been created for MobileMedia7-CIDE-Compliant

Name	Type	Rec's
Not_Favourites		0
Not_SMS_Transfer		0
Play_Music		0
SMS_Transfer		78
SMS_or_Copy		74
Single_Media_Mode		0
View_Photo		91
src/ubc/midp/mobilephot		91
ExpressionStatement		91
IfStatement		91
InsCre: MediaListSc METHO...		2

Recommendation Manager

Name	T	Value	Reasons	S...	> Val...	Range	Views
SimNam: albumname	7	1.0	TC:Check Accesses...	1		4392-4...	7
VaDeFa: imageData	4	0.76	TPF:Text Match(1), ...	4		2789-2...	0
SimNam: imageDat	7	0.75	GR:accessed by	1		3001-3...	0
SimNam: imageDat	7	0.75	GR:accessed by	1		4234-4...	0
SimNam: imageDat	7	0.37	GR:accessed by	1		4381-4...	1
View_Photo	7	0.05	TPF:Text Match	1		0-00	0
MetDec: addMediaData	2	0.55	TC:Check Decl.(1), ...	3	SMS_o...	3203-3...	0
TypDec: PhotoViewScr	1	0.55	GR:declared by(5), ...	10	SMS_o...	390-39...	0
ImpDec: javax.microed	5	0.5	GR:accessed by(1)	1		198-19...	0
ImpDec: ubc.midp.mol	5	0.5	GR:accessed by(2)	2		1077-1...	0
VaDeFa: nextcontroller	4	0.5	GR:referenced by(1...	3		15107-...	0

Domain Knowledge

Exclude code from mutually exclusive features
Include code from implied features

Experience

Qualitative

HyperSQL

ArgoUML

Quantitative

Prevayler

MobileMedia

Lampiro

Sudoku

Project	Feature	Feature Size			Mining Results		
		LOC	FR	FI	IT	Recall	Prec.
Prevayler	Censor	105 (1 %)	10	5	32	100 %	41 %
	Gzip	165 (2 %)	4	4	27	100 %	18 %
	Monitor	240 (3 %)	19	8	53	100 %	42 %
MobileM.	Replication	1487 (19 %)	37	28	64	100 %	67 %
	Snapshot	263 (3 %)	29	5	47	81 %	46 %
	Copy Media	79 (2 %)	18	6	33	97 %	26 %
	Sorting	85 (2 %)	20	6	36	96 %	46 %
	Favourites	63 (1 %)	18	6	31	100 %	43 %
	SMS Transfer	714 (15 %)	26	14	44	100 %	62 %
	Music	709 (15 %)	38	16	51	99 %	59 %
	Photo	493 (11 %)	35	13	55	99 %	49 %
	Media Transfer	153 (3 %)	4	3	25	99 %	13 %
	Lampiro	Compression	5155 (12 %)	33	20	42	100 %
	TLS Encryption	86 (0 %)	13	6	24	81 %	29 %
Sudoku	Variable Size	44 (2 %)	5	4	24	100 %	29 %
	Generator	172 (9 %)	9	7	29	98 %	42 %
	Solver	445 (23 %)	40	12	46	100 %	58 %
	Undo	39 (2 %)	5	4	29	100 %	21 %
	States	171 (9 %)	26	7	43	99 %	52 %

LOC: lines of code (and percentage of feature code in project's code base);

FR: Number of distinct code fragments; FI: Number of files; IT: Number of iterations

Extraction

Transformation

Implementation

Domain
knowledge

Legacy System with
#ifdef Variability

static variability extr.

Legacy System with
Runtime Parameters

dynamic variability extr.

Legacy System
without Variability

Feature Mining

Family of Similar
Legacy Systems

Legacy 1

Legacy 2a

Legacy 2b

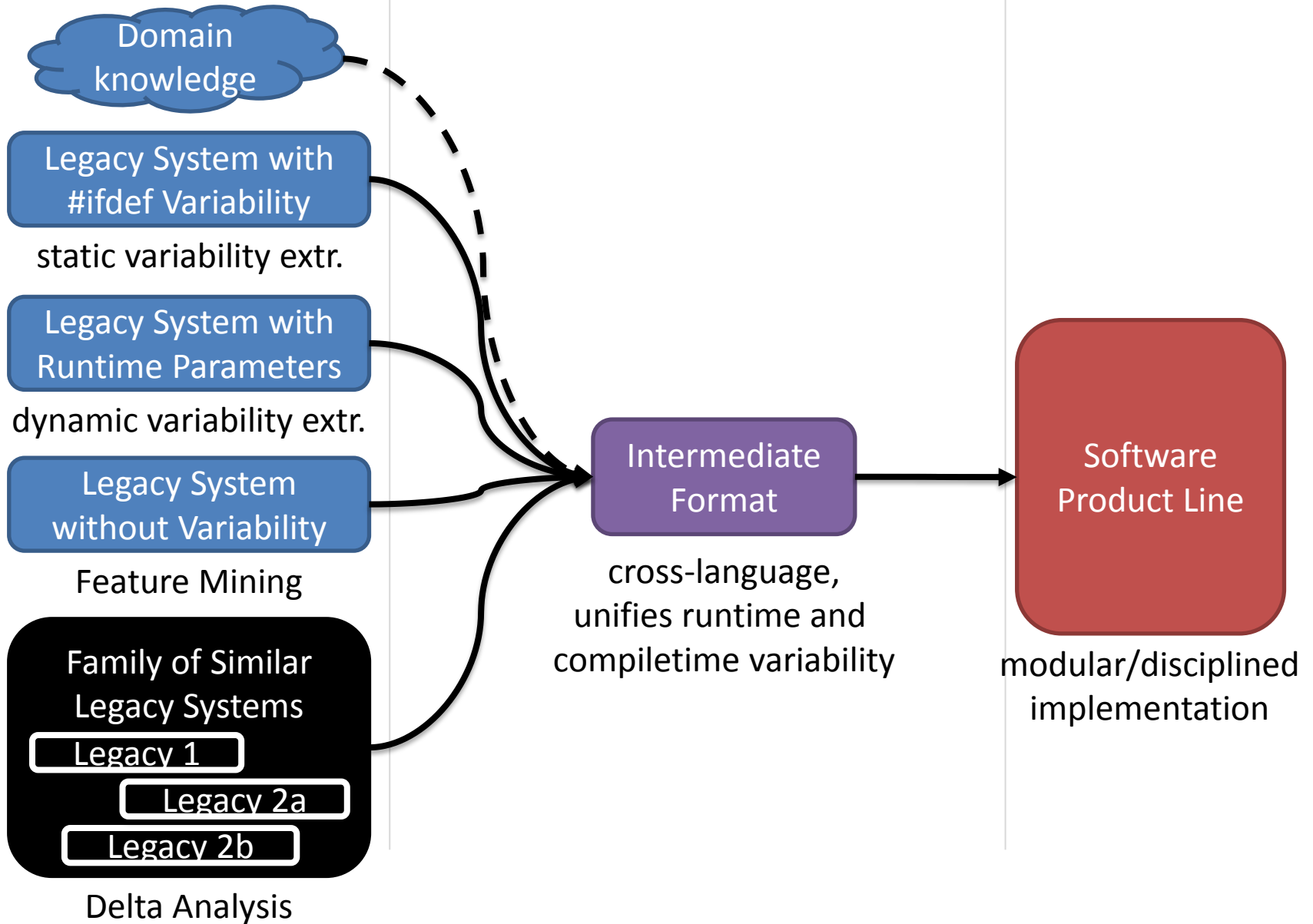
Delta Analysis

Intermediate
Format

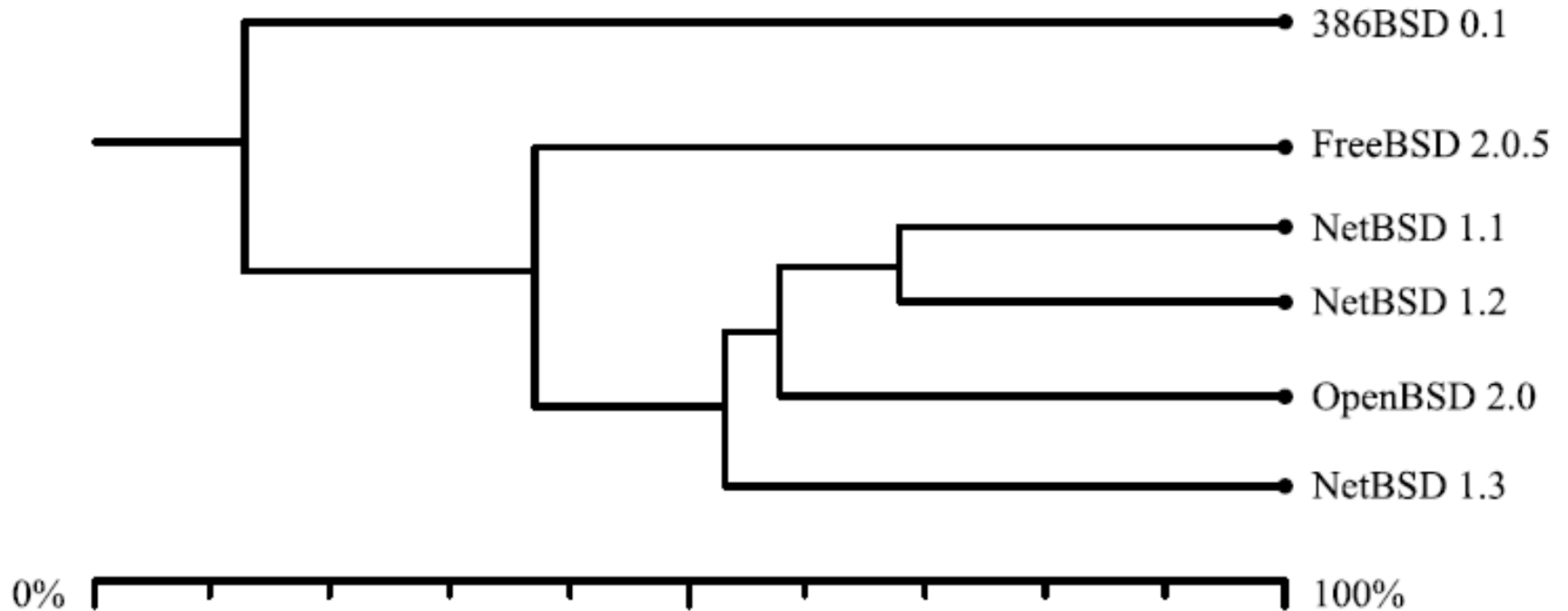
cross-language,
unifies runtime and
compiletime variability

Software
Product Line

modular/disciplined
implementation



Clone and Own



n-way diff

v1

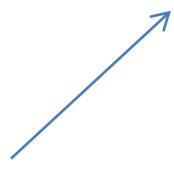
```
int a = 3;  
int b =4;  
int x = a + b;
```

v2

```
int a = 3;  
int b =2;  
int x = a + b;
```

v3

```
int a = 1;  
int b =2;  
int x = a + b;
```



A	15	20	24	10	69	
B	15	19	13	11	8	66
C	15	21	21	6	5	68
D	15	9	10	7	13	54

n-way diff

v1

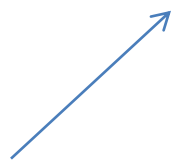
```
int a = 3;  
int b =4;  
int x = a + b;
```

v2

```
int a = 3;  
int b =2;  
int x = a + b;
```

v3

```
int a = 1;  
int b =2;  
int x = a + b;
```

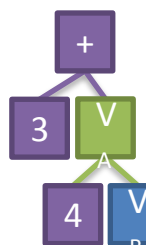


A	15	20	24	10	69	
B	15	19	13	11	8	66
C	15	21	21	6	5	68
D	15	9	10	7	13	54



```
#ifdef v1 || v2  
int a = 3;  
#else  
int a = 1  
#endif  
#ifdef v1  
int b =4;  
#else  
int b =2;  
#endif  
int x = a + b;
```

variability-aware parser



n-way diff

v1

```
int a = 3;
#ifdef A
int b =4;
#else
int b =3;
#endif
int x = a + b;
```

v2

```
int a = 3;
#ifdef B
int b =4;
#else
int b =3;
#endif
int x = a + b;
```

A	15	20	24	10	69	
B	15	19	13	11	8	66
C	15	21	21	6	5	68
D	15	9	10	7	13	54

```
int a = 3;
#ifdef v1
#ifdef A
#else
#ifdef B
#endif
int b =4;
#else
int b =3;
#endif
int x = a + b;
```

variability-aware parser

Extraction

Transformation

Implementation

Domain knowledge

Legacy System with #ifdef Variability

(variability-aware parser)

Legacy System with Runtime Parameters

(forward slicing)

Legacy System without Variability

(impact analysis)

Family of Similar Legacy Systems

Legacy 1

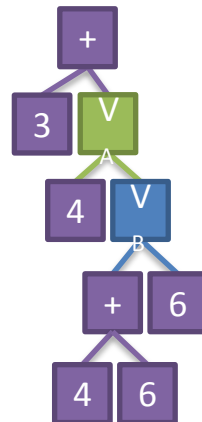
Legacy 2a

Legacy 2b

(n-way diff)

Intermediate Format

Software Product Line



4TH INTERNATIONAL WORKSHOP ON FEATURE-ORIENTED SOFTWARE DEVELOPMENT



FOSD 2012

September 24-25, 2012
Dresden, Germany



[FOSD 2009](#) [FOSD 2010](#) [FOSD 2011](#) [FOSD 2012](#)

FOSD 2012 Home

- [Abstract](#)
- [Keynotes](#)
- [Goals](#)
- [Workshop Format](#)
- [Submission](#)
- [Important Dates](#)
- [Committees](#)

Collocated Events

- [SLE 2012](#)
- [GPCE 2012](#)
- [ITSLE 2012](#)

Important Dates

- Paper submission: July 2nd, 2012
- Notification: August 8th, 2012
- Camera-ready version: August 27th, 2012
- Workshop: September 24th-25th, 2012