

Ускорение выравнивания строк на ПЛИС при помощи HaSCoL

Олег Медведев

(Ланит-терком, НИИ ИТ СПбГУ)

SECR, 02.11.2011

Множественное выравнивание

MRKMSEEEFYLFKNISSVGPWDGPQYHIAPVWAFYLQAAFMGTV
MNGTEGNNFYVPLSNRTGLVRSPFEYPQYYLAEPWQFKLLAVYM
MKQVPEFHEDFYIPIPLDINNLSAYSPFLVPQDHLGNQGIFMAM

★

FLIGFPLNAMVLVATLRYKKLRQPLNYILVNV
FFLICLGLPINGLTICTAQHKKLRQPLNFILVNL
SVFMFFIFIGGASINILTILCTIQFKKLRSHLNYILVNL

Множественное выравнивание

```
---MRKMS--EEEFYL-----FKNISSV--GPWDGPQYHIAPVWA
---MNGTE--GNNFYVP----LSNRTGLVRSPFEYPQYYLAEPWQ
---MKQVPEFHEDFYIPIPLDINNLS--AYSPFLVPQDHLGNQGI
      *               *               *               *      **
```

```
-----FYLQAAFMGTVFLIGFPLNAMVLVATLRYKKLRQPLNYILVNV
-----FKLLAVYMFFLICLGLPINGLTICTAQHKKLRQPLNFILVNL
-----FMAMSVFMFFIFIGGASINILTILCTIQFKKLRSHLNYILVNL
      *               *               *               *      ***** ** *****
```

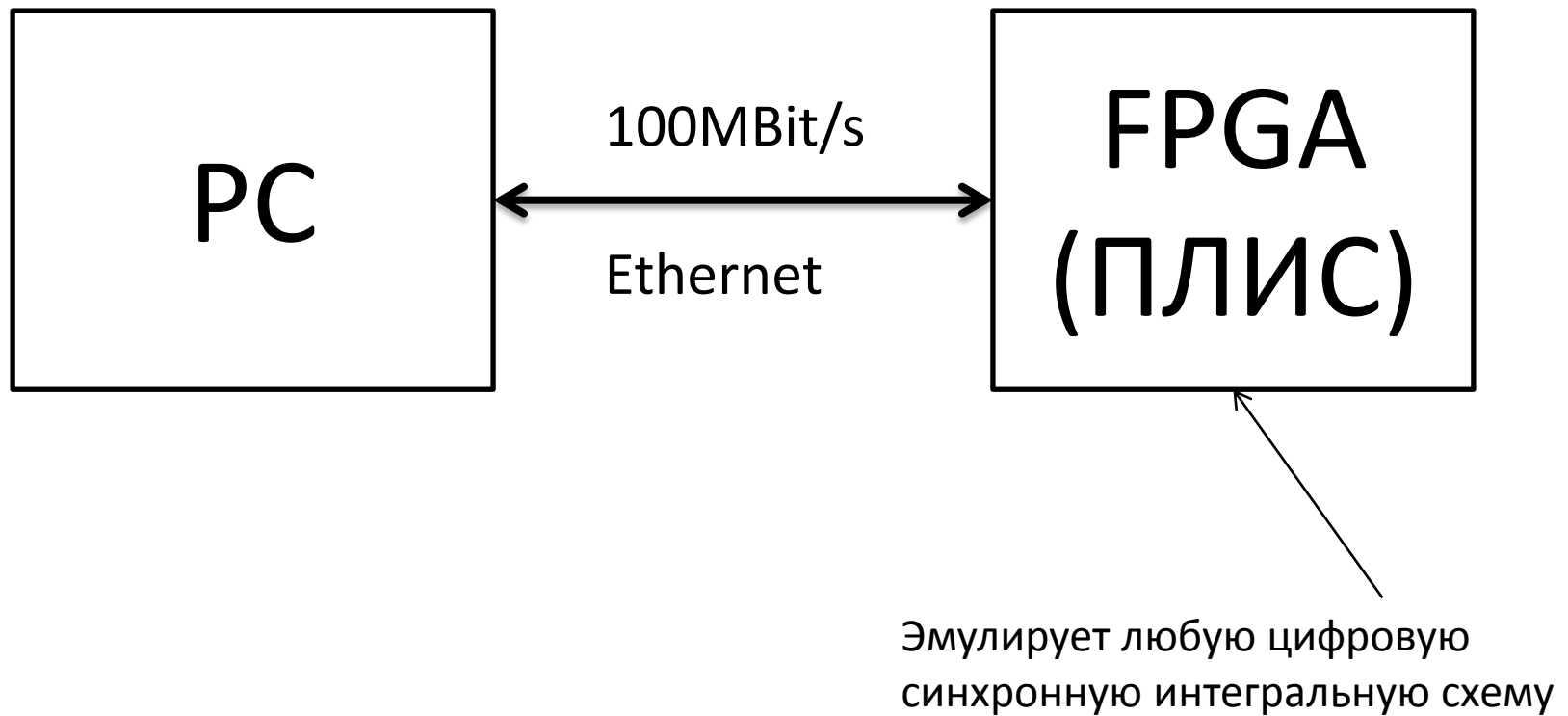
Множественное выравнивание

- Задача формализуется несколькими способами
- В любом случае точно решается экспоненциальными алгоритмами
- Поэтому на самом деле решается приближенно
- Например, при помощи open source средства MAFFT

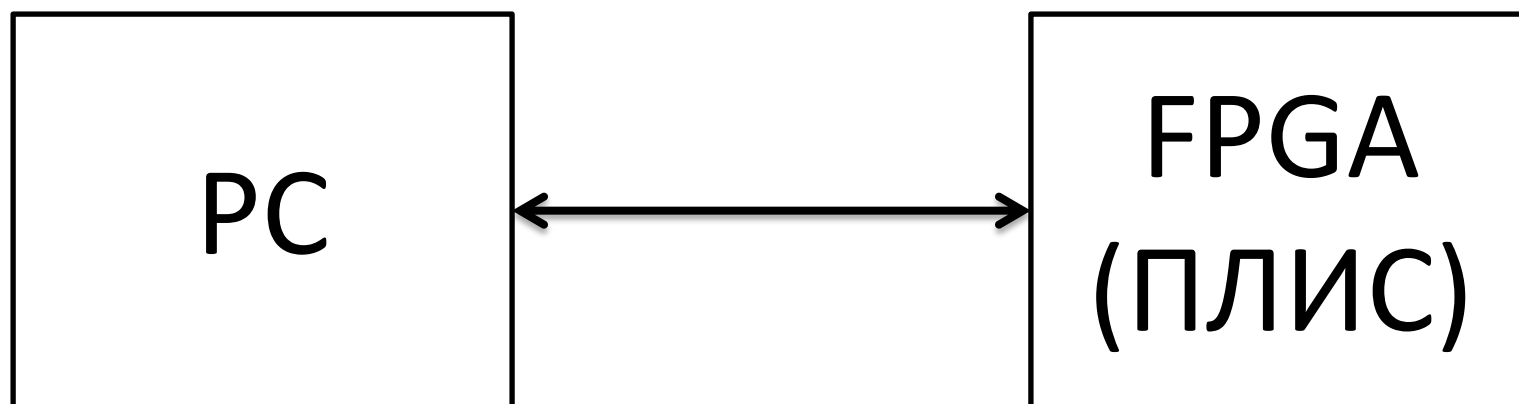
Производительность MAFFT

- Не очень хорошая (для самого «точного» варианта)
- При этом 90% времени тратится в процедуре выравнивания двух строк
- Её было решено ускорить аппаратно

Общий вид системы



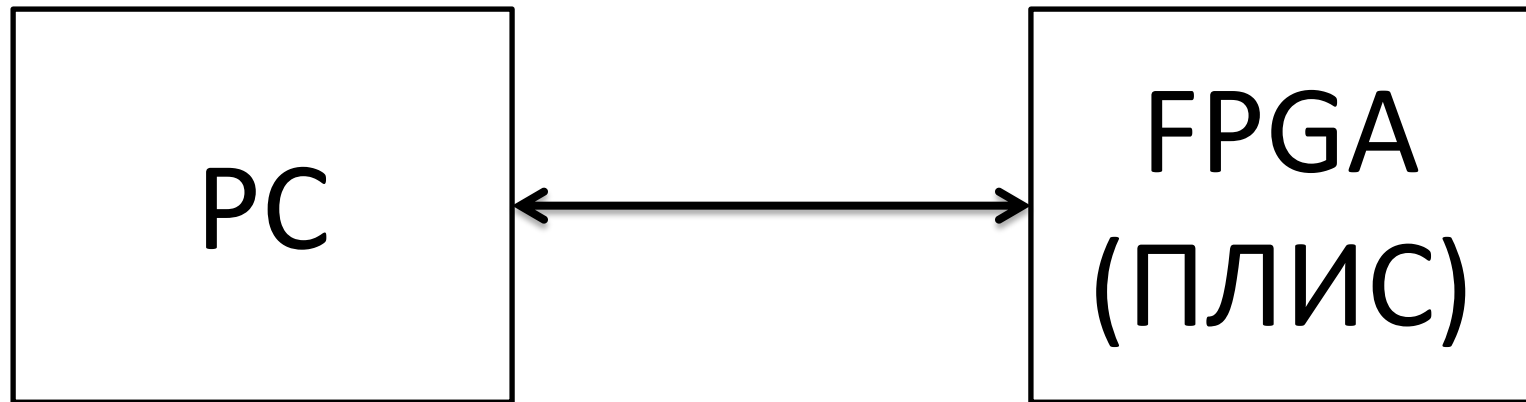
Боремся за 10х и более ускорение



Intel Core Duo 1.8GHz,
оптимизированный код
на Си

Virtex5 xc5vlx50t-1
код на HaSCoL, написанный
за 10 дней.
Выравнивание *двух строк*
Работает в 10 раз быстрее

В чем сложность подхода?



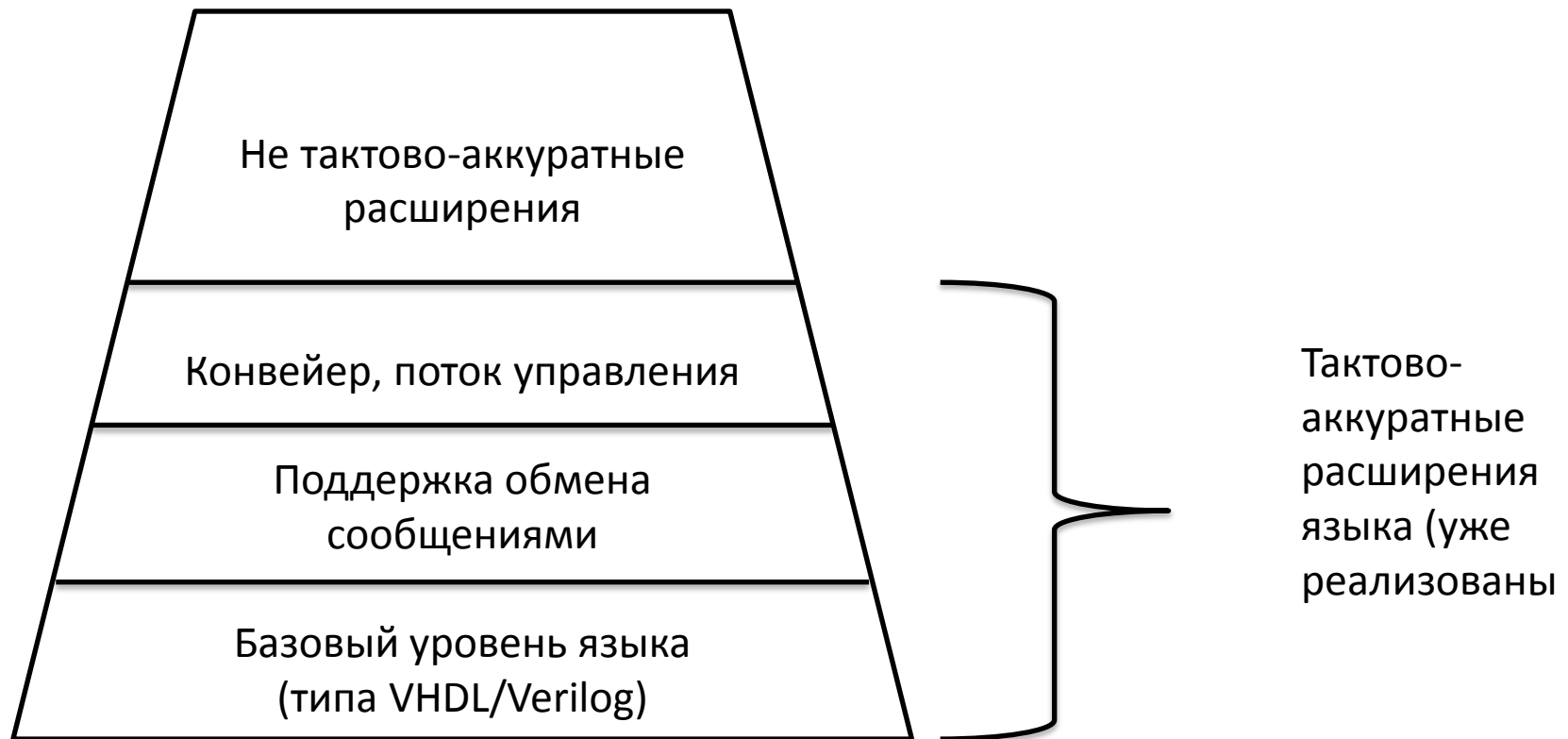
Программируется на
одном из множества
языков

Программируется на
низкоуровневых
VHDL/Verilog, либо
средствами HLS

Средства high-level synthesis

- Синтез из Си-подобных языков, либо подмножеств Си (CatapultC, HandelC, SpecC, MitrionC, ...)
- Попытки предложить свой язык (Bluespec, наш HaSCoL)

HaSCoL – набор вложенных расширяемых языков



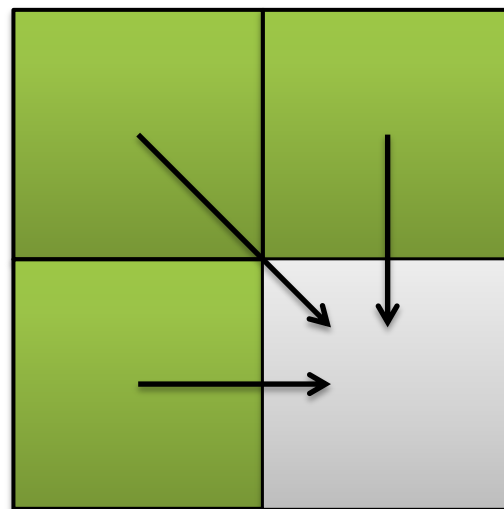
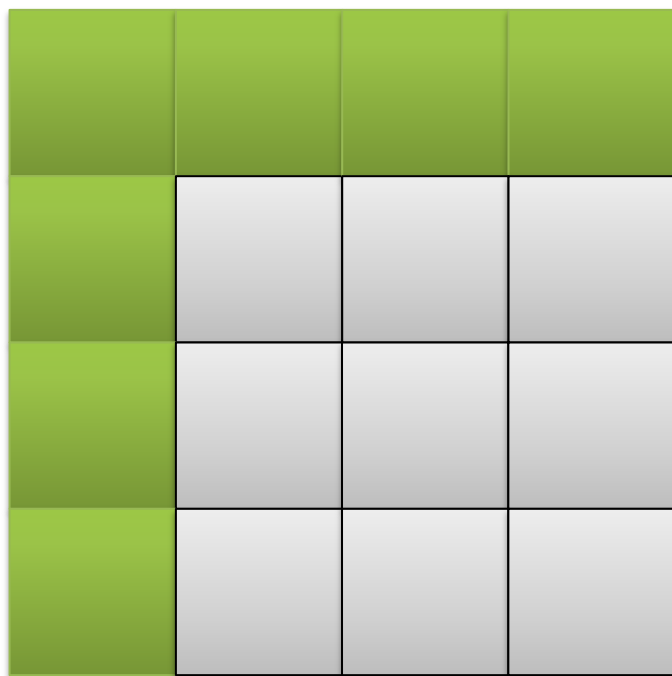
Целевая функция для выравнивания двух строк

- Даны (как-то выравненные) строки , над алфавитом
- Дана функция
- Целевая функция такова:

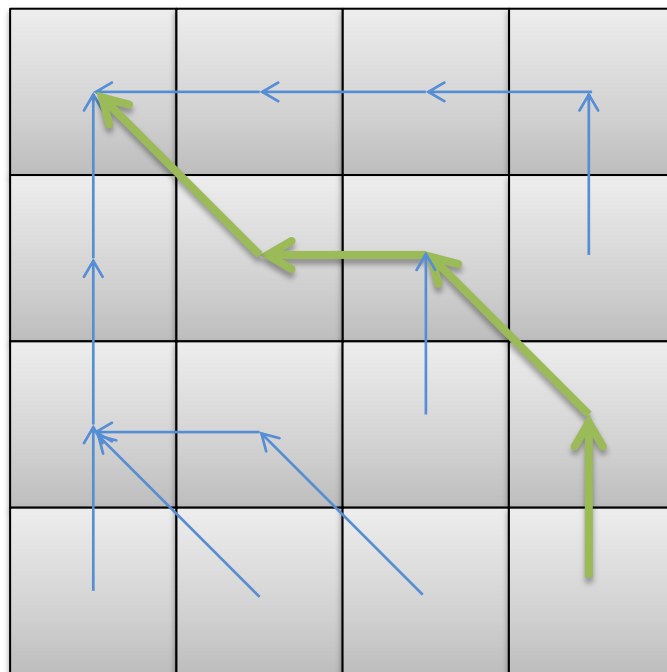
()

[] []

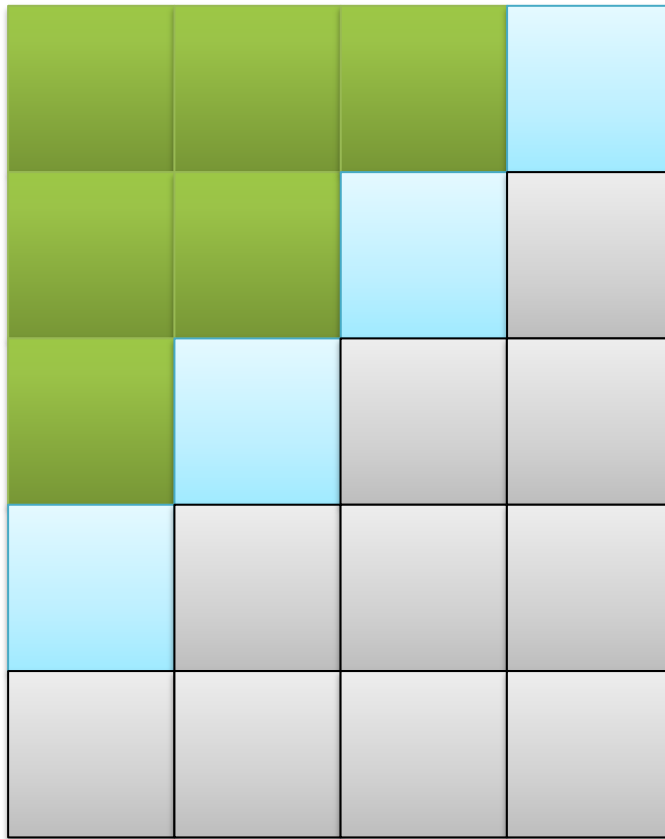
Примерное устройство решения



Матрица возвратов



Примерное устройство решения: конвейер вычислителей



Один вычислитель -- реализация

```
data previousValue : tWeight;
data previousLeftValue : tWeight;

in nextLeftValue(nextLeftValue : tWeight) {

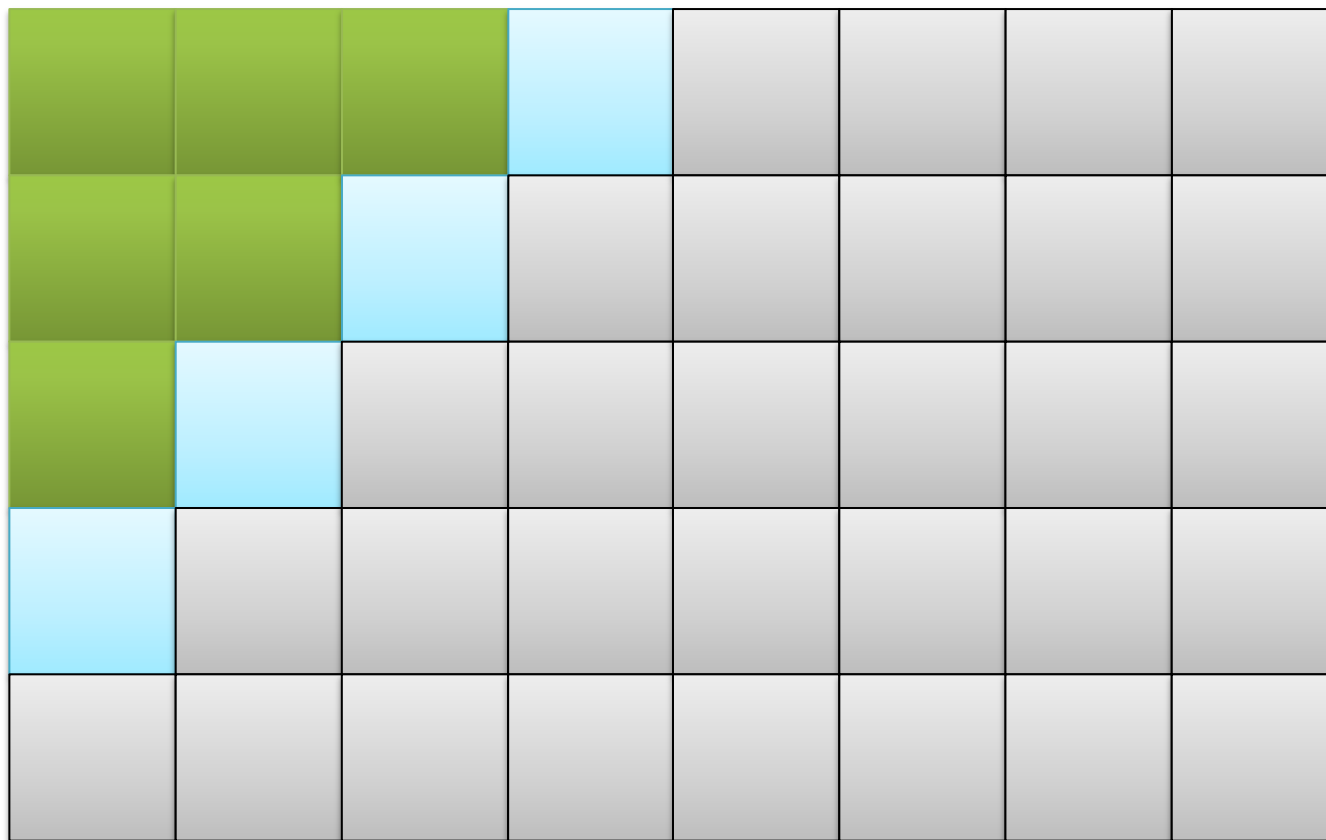
    nextOurValue = someExpression(value,
                                   previousValue, previousLeftValue)
| previousOurValue := nextOurValue
| previousLeftValue := nextLeftValue

;

    send propagateToTheRight(nextOurValue)

}
```

Сложность 1: короткий конвейер



Соответствующий управляющий код

```
while colPos < s2len do
  inform initialCol(colPos, 0)
  | colPos = colPos + AlignerSize
  | reccoldiv := reccoldiv + 1 ;
wait(AlignerSize);
let lastCol = colPos >= s2len in {
  rowPos = (0 : int(16)) | lastCol = lastCol
  | rowWritePos := 0
  | skipHead := if lastCol then
    sllenModAlignerSizeR else 0 fi
};
while rowPos <= sllen + (
  if lastCol
    then (ext(sllenModAlignerSizeR, 16) : int(16))
    else (0 : int(16)) fi
) do
  rowPos = rowPos + 1
  | inform pushRowSymb(rowPos)
done;
wait(AlignerSize)
done;
```

Сложность 2

- Необходимость делать обратный проход по матрице, перевычисляя вертикальные полосы в обратном порядке, потому что матрицу возвратов сохранить невозможно
- Эту сложность мы преодолеваем примерно таким же образом

Еще раз результаты

- Примерно за 10 дней на HaSCoL реализовано устройство вычисления попарного выравнивания
- Оно протестировано на нескольких десятках тысяч тестов в симуляции
- Синтезировано для FPGA Virtex-5 xc5v1x50t-1, по результатам синтеза получена оценка: устройство будет работать примерно в 10 раз быстрее программной реализации попарного выравнивания

Контакты, ссылки

- Язык называется HaSCoL
- Скачать тулы, статьи про HaSCoL
<http://oops.math.spbu.ru/projects/coolkit>
- GNU код на HaSCoL (разное – магистры
ГУАП, студенты, аспирант, я сам)
<http://code.google.com/p/hascolplayground>
- Oleg.Medvedev@lanit-tercom.com