

Real-Time Primer Design for DNA Chips

H. Simmler and H. Singpiel
Acconovis GmbH
Lindenhofstr. 42-44
68163 Mannheim, Germany
eMail: simmler@acconovis.com

R. Männer
Universität Mannheim
B6, 23-29
68131 Mannheim, Germany
maenner@ti.uni-mannheim.de

Abstract

The design of PCR or DNA Chip experiments is a time consuming process where bioinformatics is extensively used. The selection of the primers, which are immobilized on the DNA chip, requires a complex algorithm. Based on several parameters an optimized set of primers is automatically detected for a given gene sequence.

This paper describes a parallel architecture which performs the optimization of the primer selection on a hardware accelerator. In contrast to the pure software approach, the parallel architecture gains a speedup of factor 500 using a PCI based hardware accelerator. This approach allows an optimization of a specified primer set in real-time.

1 Introduction

Both, the amplification of DNA sequences using polymerase chain reaction (PCR) and the massive parallel analysis of genes in biological cells using DNA chips (or DNA arrays) have a great impact on modern biological research.

PCR is used to amplify a particular DNA fragment called target sequence. In general, a forward and a reverse primer is generated. The target sequence, located between the two primers, is duplicated using a complex process protocol [1].

DNA chips are used to analyse a large number of genes in parallel. This provides an insight view into cells or can improve the search for gene defects in a particular genome. The DNA chips perform up to 500.000 experiments in parallel and enable the researcher to monitor the whole genome on a single chip at the same time [2, 3].

Although these two applications have different aims – amplification and analysis – both techniques make use of primers. Formally, primers are considered as strings that represent a DNA sequence. This DNA sequence consists of four bases represented by the letters $\{A, G, T, C\}$. The start of the DNA sequence is denoted by 5' end and the termination is denoted 3' end [4].

Prior to the biological experiments, either PCR or DNA chips, primers have to be designed and synthesized. In general, primer design is based on several criteria that extend beyond string matching. Typical criteria used for the design are the exact string match, the primer length, the melting temperature, the salt concentration for the experiment and the hybridization effects that have to be taken into account for the selected primers.

PCR experiments need only a few different primers whereas several thousand different primers are needed for a DNA chip. The complete processing time for an optimal primer set can take hours taking the various criteria into account.

Preparing a DNA experiment can be described as a workflow consisting of three steps.

1. Define the genes that have to be analysed.
2. Design the optimum primer for the gene.
3. Verify the primer in a macroscopic experiment.

Furthermore, the second design step is separated into the computation of the primer sets and a database comparison with each primer. The database check compares the selected primers against the genome database to avoid a “false positive” signal that is not generated by the specified gene. This paper concentrates on the design of the primers.

In section two the basics of DNA chips are described. Section three specify various parameters that are used to select the optimal primers. The computation steps performed to select these optimal primers are described in section four. The fifth section shows the idea of the parallel architecture whereas its implementation is described in section six. The results achieved with the parallel architecture are listed in section seven. The final section provides some conclusions and further applications of the parallel architecture.

2 DNA Chips

2.1 Experiments

It is believed that thousands of genes and especially their interactions are responsible for the mystery of life. Before DNA chips were available, researchers were able to look at only a few genes at the same time.

Nowadays, DNA chips provide a complete set of biological experiments on one chip that can be performed simultaneously with one single probe. This enables researchers to have a complete look at a biological cell so that gene interactions or gene defects can be analysed within a short time. The main application fields for DNA chips are gene expression analysis, single polymorphism detection (SNPs), medical diagnostics, gene discovery, drug discovery and toxicological studies. For more information see [5].

2.2 DNA Chip Design

A DNA chip is separated into a matrix of spots. The amount of spots can vary from low density chips with 96 spots up to high density chips with 500.000 spots on one chip. Each spot on a DNA chip contains primers with a unique coding. The primers are immobilized on the spots. As shown in Figure 1 the interesting gene sequence¹ binds to the primer. Each primer base binds the corresponding position on the gene sequence. This prevents the gene sequence from swill during the washing process. A fluorescence marker is attached to the probe sequence so that the found sequence can be read out by a DNA chip reader.

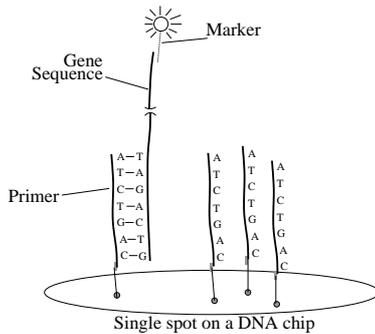


Figure 1. Sample spot on a DNA chip.

The used primers vary depending on the chip and the experiments. Usually, the primers are between 20 and 100 bases long and they are manufactured synthetically. See [3] for more information.

¹This can be a DNA sequence or a product generated by the gene.

3 Primer Design

A biological experiment needs a complete primer set that has to be designed. Choosing the optimal primers for a given target sequence or gene requires the evaluation and comparison of several partly independent parameters against a set of ideal values.

The user defines the target sequence and specifies a set of ideal parameter values. Each parameter value consists of an ideal value and a range indicating valid parameter results. The minimum and maximum values are used for filtering. They reduce the amount of primers that have to be analysed. A quality score is computed to select the optimum primer. This score is defined as the sum of all distances between the parameter result to the ideal values.

3.1 Hybridization conditions

The parameters taken into account for selecting each optimal primer set have major influence on the quality of the hybridization process where the primers react with the genes or target sequences. The condition is defined by the parameters which are described in the following paragraphs.

Primer length The primer length defines the amount of bases that build the biological primer. Primarily this length defines the selectivity [6] of the primer. Secondly it has an influence on the melting temperature and the hybridization effects. The primer length is used to generate the primers from the specified sequence windows that are evaluated. See section 4 for more details.

Melting temperature The primer and the searched gene sequence correspond to each other so that each base can bind to its counterpart. The melting temperature is the temperature at which the bonds between primer and gene dissolves. This is an important parameter for the PCR and also for the DNA chips to avoid bindings that use only a fraction of the available primer and cause a “false positive” signal.

The melting temperature for a given primer $p = (p_1, \dots, p_n)$ is calculated with a prominent approximation for the melting temperature [7, 8, 9, 10]. The formula is:

$$T_{m,1}(p) = \frac{\Delta H(p)}{\Delta S(p) + R * \ln(\frac{\gamma}{4})} + T_0 + t \quad (1)$$

where $R = 1.987(cal/^{\circ}C * mol)$ is the molar gas constant, $\gamma = 50 * 10^{-9}$ is the molar concentration of the primer in its solution, $T_0 = -237.15^{\circ}C$, and $t = -21.6^{\circ}C$ is an empirical temperature correction. The value t may depend upon the ion concentration and other unknown factors. The enthalpy $\Delta H(p)$ and the entropy $\Delta S(p)$ of

the primer p are computed according to the nearest neighbor schemata [9] $\Delta H(p) = \sum_{i=1}^{n-1} \Delta H(p_i, p_{i+1})$ and $\Delta S(p) = \sum_{i=1}^{n-1} \Delta S(p_i, p_{i+1})$ where enthalpy and entropy of a string consist of two bases. The used values for the base combinations are listed in the following Table 1. The

Nearest neighbor thermodynamics		
(p_i, p_{i+1})	$\Delta H(p_i, p_{i+1})$	$\Delta S(p_i, p_{i+1})$
AA or TT	9.1	24.0
AT	8.6	23.9
TA	6.0	16.9
CA or TG	5.8	12.9
GT or AC	6.5	17.3
CT or AG	7.8	20.8
GA or TC	5.6	13.5
CG	11.9	27.8
GC	11.1	26.7
GG or CC	11.0	26.6

Table 1. Nearest neighbor thermodynamics values [11]

values in Table 1 refer to the energy required to disrupt the hydrogen bonds of a single base pair of a paired chain. It is assumed to be influenced by neighboring bases. More details can be seen in [11].

GC content Chemically, hydrogen bonds between the bases of the primer and the gene are responsible for a stable binding. G-C pairs form three hydrogen bonds and are more stable than A-T pairs which form only two hydrogen bonds. Thus, a high GC content results in a greater stability between primer and gene.

The GC content simply measures the amount of GC bases for the primer. The following formula is used:

$$GC(p) = \frac{\#G \text{ in } p + \#C \text{ in } p}{|p|} * 100 \quad (2)$$

Secondary structure Above, only the linear sequence, also known as the primary structure, is considered. Beside this primary structure also the secondary structure and its effects have to be taken into account.

The secondary structure considers the fact that primers are flexible and that base pairs may bind to each other generating structures [6]. The DNA double helix is one example of a secondary structure. Other important secondary structures for the primer design are primer-primer bindings and hairpin, bulge or internal loops.

The secondary structure is an important criterion for the selection of the primer, because a hybridization between base pairs can disable the primer for the actual hybridization. Interaction between primers either for PCR or for DNA chips must be avoided to conserve the maximum sensitivity of the primer and the spot on the DNA chip.

The following four paragraphs show the criteria that are used for the detection of secondary structure effects. The

calculation compares the primer sequences and examines the possibility of a hybridization to itself or to another primer. All calculations use the same basic score function for comparing a base pair. This score function is defined as

$$s(p_i, q_j) = \begin{cases} 2, & \text{if } \{p_i, q_j\} = \{A, T\} \\ 4, & \text{if } \{p_i, q_j\} = \{C, G\} \\ 0, & \text{else} \end{cases} \quad (3)$$

where $p = (p_1, \dots, p_n)$ and $q = (q_1, \dots, q_m)$ are the two primers that are analysed for a possible secondary structure.

Self Annealing The self annealing (SA) calculation estimates the possibility of an unintended hybridization to the primer itself. Therefore, the SA score indicates the probability of generating hairpin and internal loops.

The calculation is done using the original primer and an opposite version of the same primer. The following example shows these two primers.

original primer 5' - TTCGTACGAAC - 3'
opposite primer 3' - CAAGCATGCTT - 5'

Both primers are defined as

original primer $p = \{p_1, \dots, p_n\}$
opposite primer $q = \{q_1, \dots, q_m\}$

The calculation of the SA score starts with the left shifted opposite primer, where only one overlapping position with the original primer exists. It compares each single overlapping position using the score function 3 and accumulates each single score values to an alignment score. After computing this alignment, the opposite primer is shifted one position to the right and the new alignment score is calculated. This will continue until there is only one overlapping position at the right end of the original primer. The maximum of all alignment scores is used as the SA score for the given primer. The complete function is defined as

$$SA(p, q) = \max_{k=-(n-1), \dots, m-1} \sum_{i=1}^n s(p_i, q_{i+k}) \quad (4)$$

The SA score calculation requires the evaluation of all possible alignments. The amount of alignments k_{SA} depend on the primer length and can be calculated using the formula $k_{SA} = (n-1) * (m-1) + 1$.

The example in Table 2 shows the calculation of the SA score for the primer $p = GATTA$. The table shows the alignments and the resulting SA scores.

Self End Annealing The self end annealing (SEA) calculation is very similar to the SA calculation but it considers only those alignments where the 3' end of the original primer belongs to the overlapping region. Furthermore, the

SHIFT	ALIGNMENT	SA SCORE
	5' -GATTA-3'	
-(n-1)	3' -ATTAG-5'	0
	5' -GATTA-3'	
-(n-2)	3' -ATTAG-5'	0
	5' -GATTA-3'	
-(n-3)	3' -ATTAG-5'	0
	5' -GATTA-3'	
-(n-4)	 3' -ATTAG-5'	4
	5' -GATTA-3'	
0	 3' -ATTAG-5'	4
	5' -GATTA-3'	
(n-4)	3' -ATTAG-5'	0
	5' -GATTA-3'	
(n-3)	 3' -ATTAG-5'	4
	5' -GATTA-3'	
(n-2)	 3' -ATTAG-5'	4
	5' -GATTA-3'	
(n-1)	3' -ATTAG-5'	0

Table 2. Self annealing calculation for the example.

SEA score is accumulated only for these overlaps which are continuous. Therefore, the SEA score evaluates the probability of generating hairpin loops or other primer-primer interaction that start from the 3' end and are continuous.

Because only these alignments are considered where the 3' end is involved there are less alignments that have to be evaluated. The amount is equal to the primer length.

The calculation of the SEA score starts in the shift position where all bases of the original and the opposite primer overlap. All overlapping positions starting from the 3' end of the original primer are accumulated using the score function 3. In case that the base pair does not match, the accumulation is aborted. After the alignment score is computed, the opposite primer is shifted to the right by one position and the new alignment score is computed. This process continues until only one position at the 3' end of the original primer overlaps. The SEA score is the maximum of all alignment scores.

The primer from the SA example is taken to show the computation of the SEA score in Table 3. Altogether, 5 alignments have to be evaluated for the primer $p = GATTA$ to achieve the SEA score. These alignments and the resulting scores are shown.

In case that a primer binds to itself or to a primer with the same sequence a secondary structure is generated. Secondary structures are also build when several primers are combined for a PCR experiment. Several primers are mixed

SHIFT	ALIGNMENT	SEA SCORE
	5' -GATTA-3'	
0	3' -ATTAG-5'	0
	5' -GATTA-3'	
(n-4)	3' -ATTAG-5'	0
	5' -GATTA-3'	
(n-3)	 3' -ATTAG-5'	2
	5' -GATTA-3'	
(n-2)	 3' -ATTAG-5'	4
	5' -GATTA-3'	
(n-1)	3' -ATTAG-5'	0

Table 3. Self end annealing calculation for the example.

in one tube where each primer can bind to each other and inhibit the reaction with the target sequence or the gene. Therefore, the effects with different primers must be considered when the optimal primers are selected.

Pair Annealing The pair annealing (PA) calculation takes the interaction of different primers into account and calculates all possible primer pairs. The models and the computing process is similar to the SA tests. Each primer pair is processed using the formula 4.

For PCR the PA scores are calculated for each primer pair that is considered to work together. On DNA chips the PA score is used to compare the primer from each spot against the primers from all other spots.

Pair End Annealing The pair end annealing (PEA) is similar to the SEA and evaluates all possible bindings starting from the 3' end of the primer. The model and the computing process of the SEA test are used for the PEA calculations.

This PEA scores are also calculated for each primer pair that can be combined.

3.2 Optimal Primer Selection

The various parameters that are described above are combined to a scoring vector. This scoring vector for PCR experiments is

$$s_{PCR}(p, q) = (length(p), GC(p), T_m(p), sa(p), sea(p), length(q), GC(q), T_m(q), sa(q), sea(q), pa(p, q), pea(p, q))$$

It is used to calculate a quality score for selecting the optimal primer pair that is needed for the PCR.

The designer specifies the hybridization conditions that are described in Section 3.1. These specified conditions

for the GC content, the melting temperature and the length are used together with the ideal values for the secondary structure to evaluate the optimal primers for the experiment. These ideal score vector for PCR is

$$sc_{PCR,ideal} = (length_f, GC_f, T_{m,f}, 0, 0, length_r, GC_r, T_{m,r}, 0, 0, 0, 0)$$

All ideal annealing scores are set to zero. The melting temperature $T_{m,f} = T_{m,r}$ and the GC content $GC_f = GC_r$ are set to be equal for the forward and reverse primer.

The deviation between the calculated value and the given ideal value is computed for each parameter of the scoring vector. Each deviation is then weighted using the values in Table 4.

Deviation from ideal length	$k1 = k6 = 0.5$
Deviation from ideal GC content	$k2 = k7 = 1$
Deviation from ideal temperature	$k3 = k8 = 1$
Deviation from ideal self annealing	$k4 = k9 = 0.1$
Deviation from ideal self end annealing	$k5 = k10 = 0.2$
Deviation from ideal pair annealing	$k11 = 0.1$
Deviation from ideal pair end annealing	$k12 = 0.2$

Table 4. Weights for the deviation calculation[11].

The resulting distance l_{PCR} , that is equal to the quality score, is calculated using

$$l_{PCR} = \sum_{i=1}^{12} k_i |sc(p, q)_i - (sc_{ideal})_i| \quad (5)$$

The scoring vector for the evaluation of a DNA chip primer considers less parameters, because only one primer has to be selected for each spot on the chip. The scoring vector and the ideal vector are defined to be

$$sc_{chip}(p) = (length(p), GC(p), T_m(p), sa(p), sea(p))$$

$$sc_{chip,ideal}(p) = (length_f, GC_f, T_{m,f}, 0, 0).$$

The distance l_{chip} used for the selection the optimal primer is calculated using the weights from Table 4 and the formula

$$l_{chip} = \sum_{i=1}^5 k_i |sc(p, q)_i - (sc_{ideal})_i| \quad (6)$$

4 Primer Set Computation

Since the calculation of primers for the DNA chip can be seen as a subset of the calculation of PCR primers the paper concentrates on the PCR primer calculation.

Additionally to the set of hybridization criteria two sequence regions from the target sequence or gene are specified by the designer. Figure 2 shows these regions that are

defines as the forward and the reverse window. All possible primer candidates for the forward primer are taken from the forward window and all possible reverse primer candidates are taken from the reverse window.

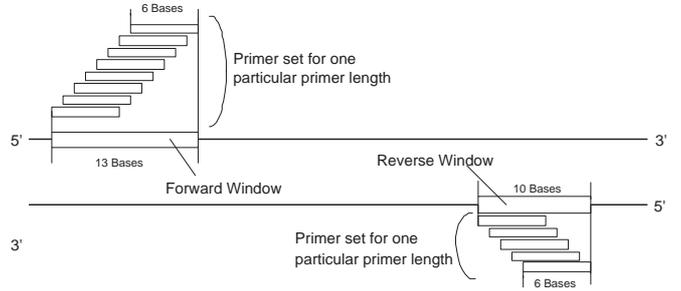


Figure 2. Primers generated from given windows.

Substrings with the length n are taken from both windows and serve as primers. The amount of these primers vary on the window length N and on the primer length n and can be calculated using $w = (N - n) + 1$. In general, the length is specified as a range and therefore the amount of possible primers for one window is

$$W = (P_1, \dots, P_w) = \sum_{l=n_{min}}^{n_{max}} (N - l) + 1 \quad (7)$$

For selecting the optimal primer pair all these primer pairs have to be computed and the distance scores l_{PCR} are calculated.

4.1 Complexity

Computing the l_{PCR} distance is a complex step that requires several operations using all given parameters. The complexity for the weighted l_{PCR} distance calculation is $O(n^2 + m^2 + nm)$. This is essentially $O(n^2)$ for computing the SA and SEA score for primer p , $O(m^2)$ for computing the SA and SEA score for primer q and $O(mn)$ for computing the PA and PEA scores.

A second step compares all primers $S = (P_1, \dots, P_w)$ from the forward window against all primers $T = (Q_1, \dots, Q_v)$ from the reverse window. This step requires $O(S^2T^2)$ distance calculations.

5 High Performance Primer Design Architecture

Calculating the quality score for one primer pair is a computationally extensive step that requires a high number of various operations. In order to keep the computation time reasonably low a parallel architecture can be used

which performs this part in a very efficient way. In our approach the overall time to compute optimal primers has only a complexity of $O(S^2T^2)$.

5.1 Architecture Overview

The calculation of the quality score consists of two main steps which can be seen in the upper part of Figure 3.

The first step computes several parameters for the given sequence pair. These parameters are the length, the GC content, the melting temperature, the SA, SEA, PA and PEA value. With these results the second step calculates a distance value. The distance is multiplied with the weights to achieve the quality score that is further used for selecting the optimal primer set for the experiments.

Due to the fact that the single parameters are independent, the parameter and the distance value calculation can be separated into two architectures which run independently. This partitioning reduce the total calculation time, because only valid primers from the first architecture are considered when the PA and PEA values are calculated. The two architectures are shown in the lower part of Figure 3.

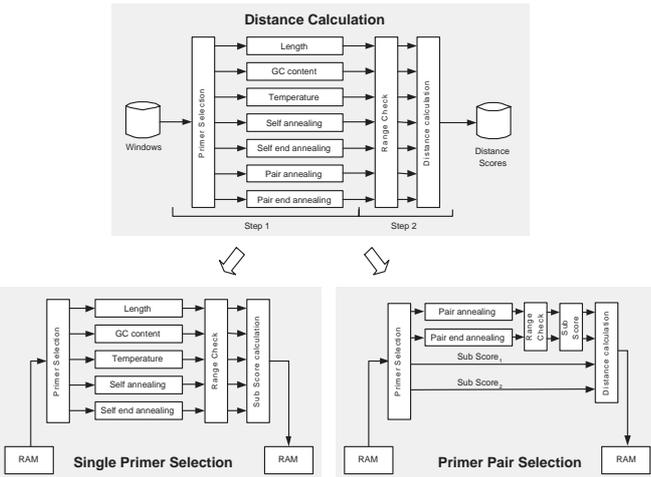


Figure 3. Architectures for the optimal primer selection.

5.2 Single Primer Selection Architecture

The first architecture performs a parallel calculation of the parameters: Length, GC content, melting temperature, SA score and SEA score. These values are calculated for each single primer defined by the length range and the window length. All forward and all reverse windows are processed, whereas each window is computed separately. The parameter results are instantly compared to the given ranges that are specified, they are multiplied with the given weights and a sub score sc_1 is calculated for each primer.

A following filter selects the single primers from each window and stores these primers using the window position, the sub score sc_1 and the length of the primer.

5.3 Primer Pair Selection Architecture

After all valid primers for each window are computed, each primer pair is evaluated to choose the optimal primer pair for the PCR. This is done using the second architecture which calculates the remaining PA and PEA values for each primer combination. The calculation of the PA and the PEA value is also done in parallel to achieve a high throughput. The results are compared to the given ranges, multiplied with the given weight and combined to the sub score sc_2 . The quality score is calculated using this sub score sc_2 and adding the sub scores of the forward primer $sc_{1,f}$ and the reverse primer $sc_{1,r}$. A filter selects the best M primer pairs and stores the quality score together with the involved primers.

6 Implementation

The two previously described architectures have been implemented on a novel computing architecture called hardware accelerator. This implementation reduces the complexity for the distance calculation from $O(n^2 + m^2 + nm)$ to $O(1)$. The implementation of the GC content, the temperature and the annealing score calculation is described in this section.

6.1 GC Content

The GC content shows the percentage of G or C bases in the primer. It is implemented as a counter followed by a divider.

Primers that will be evaluated are taken from the specified window. This primer selection shown in Figure 4 can be described as a sliding window. For each new primer length

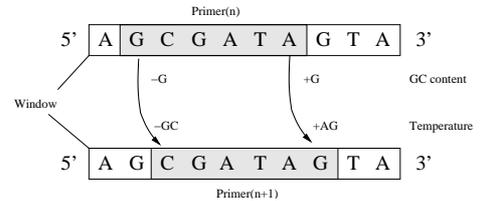


Figure 4. Primer sliding along a window.

the window starts at the 5' end and moves towards the 3' end. With every new primer the 5' base is removed and in exchange a new 3' base is attached. This circumstance is used for the implementation and makes it possible to count the GC content for a new primer with only one instruction.

In case a C or G base is added and a A or T base is removed the CG counter is incremented. If the C or G base is removed and a A or T base is added the GC counter is decremented. All other combinations have no effect on the counter.

The example in Figure 4 shows that a G base is removed and a new G base is added. Therefore, the GC content is not changed.

6.2 Melting Temperature

The same sliding window technique is used for calculating the melting temperature. The example in Figure 4 shows that a GC pair is removed and a AG pair has to be added to ΔS and ΔH . The entropy and enthalpy values are stored in a lookup table and the difference estimated through the sliding window is added to the present values. Afterwards the division is performed and the T_0 and the t constants are added.

The operations needed to compute the temperature are arranged in a pipeline, so that the temperature of a new primer can be computed with every clock cycle.

6.3 Annealing Matrix

As shown in Section 4.1 the time needed to compute the distance is dominated by calculating the annealing values. The novel annealing matrix shown in Figure 5 allows to process a new annealing value with every clock cycle. Together with the other parallel working blocks in both architectures a single primer or primer pair can be calculated every clock cycle.

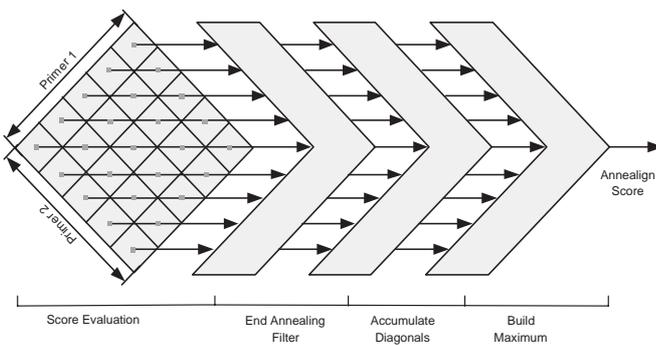


Figure 5. Schematic view of the annealing matrix.

The parallel architecture for computing the annealing scores base in a rotated matrix shown in Figure 5 on the left side. For either self or pair annealing the primers are applied to the matrix from the left side. All possible alignments have a different overlapping length. They are equal to the diagonals of the matrix.

Each single matrix cell compares the attached primer bases and performs the score function 3. The cells generate also a signal for the end annealing calculation indicating that the mismatch occurred.

The SA and the PA values are accumulated along each diagonal for all possible alignments. After the diagonal scores are computed, the maximum value of all the alignment results is generated. This output is the SA/PA score. The accumulation and the maximum building are performed in separated units using pipelines.

The SEA and the PEA scores are computed using the same blocks. The abort of the calculation on a mismatch is implemented through an additional disable signal that permits an accumulation. These signals are generated in a separate block and applied to the accumulation unit.

This annealing architecture described above allows to process a primer pair every second clock cycle. The output are the two annealing scores, either SA and SEA or PA and PEA. The current matrix size is 16×16 resulting in 768 parallel executed operations performed in only two clock cycles.

6.4 Hardware Accelerator

The annealing matrix is mapped together with other computing blocks and additional core logic onto a hardware accelerator. The current implementation use a PCI based hardware accelerator that is controlled by a standard PC. In general a suitable hardware accelerator [12] can be used to execute the time extensive parts.

The central computing core used on the hardware accelerators is a field programmable gate array (FPGA). This FPGA consists of a highly parallel structure of processing elements that can be configured for a specific application. In general, high performance is achieved using a specific parallel implementation for the application.

6.5 Application Features

Currently, the implemented architecture on the hardware accelerator can process a total amount of 512 primer pairs. The implemented annealing matrix size is 16×16 . These features were chosen for a specific task and can be easily modified and extended to the user specific application.

7 Results

The overall time needed to process a set of 512 primer pairs is measured on a 500MHz Celeron processor and is compared to the execution time using the same Celeron system with an integrated hardware accelerator. The software version performs the identical functionality as the parallel architecture.

The version supported by an hardware accelerator executes the two implemented architectures shown in Figure 3. These are implemented for a PCI based hardware accelerator and run at 50 MHz. The following Table 5 shows the times measured for both applications. Based on these mea-

PURE SOFTWARE	HARDWARE ACCELERATED
1920 sec	3,41 sec

Table 5. Execution times for 512 primer pairs

surements for the entire application a speedup factor greater than 500 is achieved by using the hardware accelerator. This remarkable speedup is also valid considering state of the art processors because the implemented matrix can be easily expanded in size using modern hardware accelerators.

8 Conclusion

This paper shows a solution that accelerates the execution time of a primer design application by a factor greater than 500. The accelerated application selects a handful optimal primers that can be used for polymerase chain reaction (PCR) as well as for DNA chip experiments. The optimal primers were selected from a defined set of possible primers specified through the target sequence or gene sequence and the length range. Various parameters are computed for all these candidates. The optimal primers are selected using a quality score that is calculated upon these parameters.

It is shown that the calculation of the quality score has a complexity $O(n^2 + m^2 + mn)$. This complexity is dominated by the calculation of the annealing values. Therefore, a matrix based parallel architecture is used to process the four different annealing value. Additionally, the other parameter are also processed in parallel. Thus, this novel computing architecture process a given primer pair every clock cycle and reduce the complexity to $O(1)$. This improvement reduce the overall complexity to $O(S^2T^2)$.

The entire distance calculation process is split into two architectures which are mapped onto a hardware accelerator. The execution of both architectures at 50 MHz lead to a total speedup factor of more than 500 compared to a 500MHz Celeron processor.

Using a parallel architecture executed on a hardware accelerator provides very high performance for this primer design application. Other applications like contig assembly or a precise primer based database search can also gain using the annealing matrix. In general, hardware accelerators can be used universally to increase the performance of standard workstations by several orders of magnitude.

9 Acknowledgements

We would like to thank Prof. Dr. R. Männer from the University of Mannheim for the support and the ideas and Chandra Sarkar from Biomerns.net for contributing some specific aspects of primers, PCR and DNA chips.

The parallel architecture for computing the annealing vector is part of a filed patent.

References

- [1] Brown, T.A.(1998): Genetics: A Molecular Approach. Chapman & Hall, London.
- [2] Blanchard, A.P.(1998): Synthetic DNA Arrays. Genetic Engineering Vol.20, 111-123.
- [3] Singlair, B. (1999): Everything's Great When It Sit's on a Chip. The Scientist13[11]:18.
- [4] Matthews, H.R., Freeland, R. and Miesfeld, R.L.(1997): Biochemistry - a short course. New York.
- [5] Wallace, R.W. (1997): DNA on a chip serving up the genome for diagnostics and research. Molecular Medicine Today, Vol.3, Issue 9, 348-389.
- [6] Kaderali, L. (2001): Selecting Target Specific Probes for DNA Arrays. Universität Köln.
- [7] Rychlik, W. and Rhoads, R.E.(1989): A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing an in vitro amplification of DNA. Nucleic Acids Res., 19, 8543-8551.
- [8] Borer, P.N. et.al.(1974): Stability of ribonucleid acid double-stranded helicies. J. Mol. Biol., (86), 843-853.
- [9] Breslauer, K.J. et.al.(1986): Predicting DNA duplex stability from the base sequence. Proc.Natl.Acad.Sci.USA, 83, 3746-3750.
- [10] Freier, S.M. et.al.(1994): Improving free-energy parameters for predictions of RNA duplex stability. Proc.Natl. Acad. Sci. USA. 83, 9373-9377.
- [11] Kämpke, T. et.al.(2001): Efficient primer design algorithms. Bioinformatics Vol.17, 214-225.
- [12] Optimagic Inc. (2002): Programmable Logic Jump Station, <http://www.optimagic.com>.
- [13] Rychlik, W.(1995): Selection of primers for polymerase chain reaction. Mol.Biotechnol., 3, 129-143.
- [14] Edwards, M.C. and Gibbs, R.A.(1994): Multiplex PCR: advantages, development and applications. PCR Meth. Appl., 3, 565-575.