NESL: A Nested Data-Parallel Language
(Version 2.6)

Guy E. Blelloch
April 1993
CMU-CS-93-129

(Updated version of CMU-CS-92-103, January 1992)

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

This research was sponsored by the Avionics Laboratory, Wright Research and Development Center, Aeronautical Systems Division (AFSC), U.S. Air Force, Wright-Patterson AFB, Ohio 45433-6543 under Contract F33615-90-C-1465. ARPA Order No. 7997. The author is also supported by a Finmeccanica chair and an NSF Young Investigator award.

The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. government.
Keywords: Data-parallel, parallel algorithms, supercomputers, nested parallelism, PRAM model, parallel programming languages, collection-oriented languages.
Abstract

This report describes NESL, a strongly-typed, applicative, data-parallel language. NESL is intended to be used as a portable interface for programming a variety of parallel and vector supercomputers, and as a basis for teaching parallel algorithms. Parallelism is supplied through a simple set of data-parallel constructs based on sequences (ordered sets), including a mechanism for applying any function over the elements of a sequence in parallel and a rich set of parallel functions that manipulate sequences.

NESL fully supports nested sequences and nested parallelism—the ability to take a parallel function and apply it over multiple instances in parallel. Nested parallelism is important for implementing algorithms with complex and dynamically changing data structures, such as required in many graph and sparse matrix algorithms. NESL also provides a mechanism for calculating the asymptotic running time for a program on various parallel machine models, including the parallel random access machine (PRAM). This is useful for estimating running times of algorithms on actual machines and, when teaching algorithms, for supplying a close correspondence between the code and the theoretical complexity.

This report defines NESL and describes several examples of algorithms coded in the language. The examples include algorithms for median finding, sorting, string searching, finding prime numbers, and finding a planar convex hull. NESL currently compiles to an intermediate language called VCODE, which runs on the Cray Y-MP, Connection Machine CM-2, and Encore Multimax. For many algorithms, the current implementation gives performance close to optimized machine-specific code for these machines.

Note: This report is an updated version of CMU-CS-92-103, which described version 2.4 of the language. The most significant changes in version 2.6 are that it supports polymorphic types, has an ML-like syntax instead of a lisp-like syntax, and includes support for I/O.
1 Introduction

This report describes and defines the data-parallel language NESL. The language was designed with the following goals:

1. To support parallelism by means of a set of data-parallel constructs based on sequences. These constructs supply parallelism through (1) the ability to apply any function concurrently over each element of a sequence, and (2) a set of parallel functions that operate on sequences, such as the permute function, which permutes the order of the elements in a sequence.

2. To support complete nested parallelism. NESL fully supports nested sequences, and the ability to apply any user defined function over the elements of a sequence, even if the function is itself parallel and the elements of the sequence are themselves sequences. Nested parallelism is critical for describing both divide-and-conquer algorithms and algorithms with nested data structures [5].

3. To generate efficient code for a variety of architectures, including both SIMD and MIMD machines, with both shared and distributed memory. NESL currently generates a portable intermediate code called VCODE [7], which runs on the CRAY Y-MP, the Connection Machine CM-2, and the Encore Multimax. Various benchmark algorithms achieve very good running times on these machines [12, 6].

4. To be well suited for describing parallel algorithms, and to supply a mechanism for deriving the theoretical running time directly from the code. Each function in NESL has two complexity measures associated with it, the work and step complexities [5]. A simple equation maps these complexities to the asymptotic running time on a Parallel Random Access Machine (PRAM) Model.

NESL is a strongly-typed strict first-order functional (applicative) language. It runs within an interactive environment and is loosely based on the ML language [27]. The language is uses sequences (ordered sets) as a primitive parallel data type, and parallelism is achieved exclusively through operations on these sequences [5]. The set of sequence functions supplied by NESL was chosen based both on their usefulness on a broad variety of algorithms, and on their efficiency when implemented on parallel machines. To promote the use of parallelism, NESL supplies no serial looping constructs (although serial looping can be simulated with recursion), and supplies no data-structures that require serial access, such as lists in Lisp or ML.

NESL is the first data-parallel language whose implementation supports nested parallelism. Nested parallelism is the ability to take a parallel function and apply it over multiple instances in parallel—for example, having a parallel sorting routine, and then using it to sort several sequences concurrently. The data-parallel languages C* [31], *Lisp [24], and Fortran 90 [1] (with array extensions) support no form of nested parallelism. The parallel collections in these languages can only contain scalars or fixed sized records. There is also no means in these languages to apply a user defined function over each element of a collection. This prohibits the expression of any form of nested parallelism. The languages Connection
Machine Lisp [38], and Paralation Lisp [32] both supply nested parallel constructs, but no implementation ever supported the parallel execution of these constructs. Blelloch and Sabot implemented an experimental compiler that supported nested-parallelism for a small subset of Paralation Lisp [9], but it was deemed near impossible to extend it to the full language.

A common complaint about high-level data-parallel languages and, more generally, in the class of Collection-Oriented languages [35], such as SETL [33] and APL [22], is that it can be hard or impossible to determine approximate running times by looking at the code. As an example, the \( \beta \) primitive in CM-Lisp (a general communication primitive) is powerful enough that seemingly similar pieces of code could take very different amounts of time depending on details of the implementation of the operation and of the data structures. A similar complaint is often made about the language SETL—a language with sets as a primitive data structure. The time taken by the set operations in SETL is strongly affected by how the set is represented. This representation is chosen by the compiler.

For this reason, NESL was designed so that the built-in functions are quite simple and so that the asymptotic complexity can be derived from the code. To derive the complexity, each function in NESL has two complexity measures associated with it: the work and step complexities [5]. The work complexity represents the serial work executed by a program—the running time if executed on a serial RAM. The step complexity represents the deepest path taken by the function—the running time if executed with an unbounded number of processors. Simple composition rules can be used to combine the two complexities across expressions and, based on Brent's scheduling principle [10], the two complexities place an upper bound on the asymptotic running times for the parallel random access machine (PRAM) [16].

The current compiler translates NESL to VCODE [7], a portable intermediate language. The compiler uses a technique called flattening nested parallelism [9] to translate NESL into the much simpler flat data-parallel model supplied by VCODE. VCODE is a small stack-based language with about 100 functions all of which operate on sequences of atomic values (scalars are implemented as sequences of length 1). A VCODE interpreter has been implemented for running VCODE on the Cray Y-MP, Connection Machine CM-2, or any serial machine with a C compiler [6]. The sequence functions in this interpreter have been highly optimized [5, 14] and, for large sequences, the interpretive overhead becomes relatively small yielding high efficiencies. For the Encore Multimax Chatterjee has developed a compiler for VCODE [12, 13]. This compiler reduces both the synchronization needed among processors and the memory traffic over the shared bus. Most of the techniques used by this VCODE compiler should be applicable to any MIMD parallel machine.

The interactive NESL environment runs within Common Lisp and can be used to run VCODE on remote machines. This allows the user to run the environment, including the compiler, on a local workstation while executing interactive calls to NESL programs on the CRAY Y-MP or CM-2 (or any other workstation, if so desired). As in the Standard ML of New Jersey compiler [2], all interactive invocations are first compiled (in our case into VCODE), and then executed.

Control parallel languages that have some feature that are similar to NESL include ID [28, 3], Sisal [25], and Proteus [26]. ID and Sisal are both side-effect free and supply
operations on collections of values.

The remainder of this section discusses the use of sequences and nested parallelism in NESL, and how complexity can be derived from NESL code. Section 2 shows several examples of code, and Section 3 along with Appendix A and Appendix B defines the language. Shortcomings of NESL include the limitation to first-order functions (there is no ability to pass functions as arguments). We are currently working on a follow-up on NESL, which will be based on a more rigorous type system, and will include some support for higher-order functions.

1.1 Parallel Operations on Sequences

NESL supports parallelism through operations on sequences. A sequence is an ordered set and is specified in NESL using square brackets. For example

\[ [2, 1, 9, -3] \]

is a sequence of four integers. In NESL all elements of a sequence must be of the same type, and all sequences must be of finite length. Parallelism on sequences can be achieved in two ways: the ability to apply any function concurrently over each element of a sequence, and a set of built-in parallel functions that operate on sequences. The application of a function over a sequence is achieved using set-like notation similar to set-formers in SETL \[33\] and list-comprehensions in Miranda \[36\] and Haskell \[21\]. For example, the expression

\[ \{\text{negate}(a) \mid a \in [3, -4, -9, 5]\}; \]

\[ \Rightarrow [-3, 4, 9, -5] : \text{[int]} \]

negates each elements of the sequence \([3, -4, -9, 5]\). This construct can be read as “in parallel for each \(a\) in the sequence \([3, -4, -9, 5]\), negate \(a\)”. The symbol \(\Rightarrow\) points to the result of the expression, and the expression \([\text{int}]\) specifies the type of the result: a sequence of integers. The semantics of the notation differs from that of SETL, Miranda or Haskell in that the operation is defined to be applied in parallel. Henceforth we will refer to the notation as the apply-to-each construct. As with set comprehensions, the apply-to-each construct also provides the ability to subselect elements of a sequence: the expression

\[ \{\text{negate}(a) \mid a \in [3, -4, -9, 5] \mid a < 4\}; \]

\[ \Rightarrow [-3, 4, 9] : \text{[int]} \]

can be read as, “in parallel for each \(a\) in the sequence \([3, 4, 9, 1]\) such that \(a\) is less than 4, negate \(a\)”. The elements that remain maintain their order relative to each other. It is also possible to iterate over multiple sequences. The expression

\[ \{a + b \mid a \in [3, -4, -9, 5]; b \in [1, 2, 3, 4]\}; \]

\[ \Rightarrow [4, -2, -6, 9] : \text{[int]} \]

elementwise adds the two sequences. A full description of the apply-to-each construct is given in Section 3.2.

In NESL, any function, whether primitive or user defined, can be applied to each element of a sequence. So, for example, we could define a factorial function
<table>
<thead>
<tr>
<th>Operation</th>
<th>Description</th>
<th>Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>* dist(a,1)</td>
<td>Distribute value a to sequence of length 1.</td>
<td>S(result)</td>
</tr>
<tr>
<td>* #a</td>
<td>Return length of sequence a.</td>
<td>1</td>
</tr>
<tr>
<td>a[i]</td>
<td>Return element at position i of a.</td>
<td>S(result)</td>
</tr>
<tr>
<td>rep(v,d,i)</td>
<td>Replace element at position i of d with v.</td>
<td>S(v), S(d)</td>
</tr>
<tr>
<td>[s:e]</td>
<td>Return integer sequence from s to e.</td>
<td>(e - s)</td>
</tr>
<tr>
<td>[s:e:d]</td>
<td>Return integer sequence from s to e by d.</td>
<td>(e - s)/d</td>
</tr>
<tr>
<td>sum(a)</td>
<td>Return sum of sequence a.</td>
<td>S(a)</td>
</tr>
<tr>
<td>* @:scan(a)</td>
<td>Return scan based on operator @.</td>
<td>S(a)</td>
</tr>
<tr>
<td>count(a)</td>
<td>Count number of true flags in a.</td>
<td>S(a)</td>
</tr>
<tr>
<td>permute(a,i)</td>
<td>Permute elements of a to positions i.</td>
<td>S(a)</td>
</tr>
<tr>
<td>* d &lt;- a</td>
<td>Place elements a in d.</td>
<td>S(a), S(d)</td>
</tr>
<tr>
<td>* a -&gt; i</td>
<td>Get values from sequence a based on indices i.</td>
<td>S(i)</td>
</tr>
<tr>
<td>pack(a,f)</td>
<td>Pack sequence a based on flags f.</td>
<td>S(a)</td>
</tr>
<tr>
<td>max_index(a)</td>
<td>Return index of the maximum value.</td>
<td>S(a)</td>
</tr>
<tr>
<td>min_index(a)</td>
<td>Return index of the minimum value.</td>
<td>S(a)</td>
</tr>
<tr>
<td>a ++ b</td>
<td>Append sequences a and b.</td>
<td>S(a) + S(b)</td>
</tr>
<tr>
<td>drop(a,n)</td>
<td>Drop first n elements of sequence a.</td>
<td>S(result)</td>
</tr>
<tr>
<td>take(a,n)</td>
<td>Take first n elements of sequence a.</td>
<td>S(result)</td>
</tr>
<tr>
<td>rotate(a,n)</td>
<td>Rotate sequence a by n positions.</td>
<td>S(a)</td>
</tr>
<tr>
<td>* flatten(a)</td>
<td>Flatten nested sequence a.</td>
<td>S(a)</td>
</tr>
<tr>
<td>* partition(a,1)</td>
<td>Partition sequence a into nested sequence.</td>
<td>S(a)</td>
</tr>
<tr>
<td>split(a,f)</td>
<td>Split a into nested sequence based on flags f.</td>
<td>S(a)</td>
</tr>
<tr>
<td>bottop(a)</td>
<td>Split a into nested sequence.</td>
<td>S(a)</td>
</tr>
</tbody>
</table>

Figure 1: List of some of the sequence functions supplied by NESL. In the work column, S(v) refers to the size of the object v. The * before certain functions means that those functions are primitives. All the other functions can be built out of the primitives with at most a constant factor overhead in both work and number of steps. For @:scan the @ can be one of {plus, max, min, or, and}. All the sequence functions are described in detail in Appendix B.2. In rep and <-, the work complexity depends on whether the variable used for d is the final reference to that variable (arguments are evaluated left to right). If it is the final reference, then the complexity before the comma is used, otherwise the complexity after the comma is used.
function factorial(i) =
  if (i == 1) then 1
  else i*factorial(i-1);
⇒ factorial : int -> int

and then apply it over the elements of a sequence

{factorial(x) : x in [3,1,7]};
⇒ [6,1,5040] : [int]

In this example, the function name(arguments) = body; construct is used to define factorial. The function is of type int -> int, indicating a function that maps integers to integers. The type is inferred by the compiler.

An apply-to-each construct applies a body to each element of a sequence. We will call each such application an instance. Since there are no side effects in NESL, there is no way to communicate among the instances of an apply-to-each. An implementation can therefore execute the instances in any order it chooses without changing the result. In particular, the instances can be implemented in parallel, therefore giving the apply-to-each its parallel semantics.

In addition to the apply-to-each construct, a second way to take advantage of parallelism in NESL is through a set of sequence functions. The sequence functions operate on whole sequences and all have relatively simple parallel implementations. For example the function sum sums the elements of a sequence.

sum([2, 1, -3, 11, 5]);
⇒ 16 : int

Since addition is associative, this can be implemented on a parallel machine in logarithmic time using a tree. Another common sequence function is the permute function, which permutes a sequence based on a second sequence of indices. For example:

permute("nesl", [2,1,3,0]);
⇒ "lens" : [char]

In this case, the 4 characters of the string "nesl" (the term string is used to refer to a sequence of characters) are permuted to the indices [2, 1, 3, 0] (n → 2, e → 1, s → 3, and l → 0). The implementation of the permute function on a distributed-memory parallel machine could use its communication network and the implementation on a shared-memory machine could use an indirect write into the memory.

Table 1 lists some of the sequence functions available in NESL. A subset of the functions (the starred ones) form a complete set of primitives. These primitives, along with the scalar operations and the apply-to-each construct, are sufficient for implementing the other functions in the table with at most a constant factor increase in both the step and work...
function kth_smallest(s, k) =
    let pivot = s[s/2];
    lesser = {e in s | e < pivot} in if (k <= #lesser) then kth_smallest(lesser, k)
    else
        let greater = {e in s | e > pivot} in if (k >= s - #greater) then
            kth_smallest(greater, k - (s - #greater))
        else pivot;

Figure 2: An implementation of order statistics. The function kth_smallest returns the kth smallest element from the input sequence s.

complexities, as defined in Section 1.5. The table also lists the work complexity of each function, which will also be defined in Section 1.5.

We now consider an example of the use of sequences in NESL. The algorithm we consider solves the problem of finding the kth smallest element in a set s, using a parallel version of the quickorder algorithm [19]. Quickorder is similar to quicksort, but only calls itself recursively on either the elements lesser or greater than the pivot. The NESL code for the algorithm is shown in Figure 2. The let construct is used to bind local variables (see Section 3.2.2 for more details.). The code first binds len to the length of the input sequence s, and then extracts the middle element of s as a pivot. The algorithm then selects all the elements less than the pivot, and places them in a sequence that is bound to lesser. For example:

\[
\begin{align*}
s &= [4, 8, 2, 3, 1, 7, 2] \\
pivot &= 3 \\
\{x \in s | x < pivot\} &= [2, 1, 2]
\end{align*}
\]

After the pack, if the number of elements in the set lesser is greater than k, then the kth smallest element must belong to that set. In this case, the algorithm calls kth_smallest recursively on lesser using the same k. Otherwise, the algorithm selects the elements that are greater than the pivot, again using pack, and can similarly find if the kth element belongs in the set greater. If it does belong in greater, the algorithm calls itself recursively, but must now readjust k by subtracting off the number of elements lesser and equal to the pivot. If the kth element belongs in neither lesser nor greater, then it must be the pivot, and the algorithm returns this value.

1.2 Nested Parallelism

In NESL the elements of a sequence can be any valid data item, including sequences. This rule permits the nesting of sequences to an arbitrary depth. A nested sequence can be written as

\[\text{This is not strictly true since some of the utility functions, such as reading or writing from a file, have side effects. These functions, however, cannot be used within an apply-to-each construct.}\]
Table 1: Routines with nested parallelism. Both the inner part and the outer part can be executed in parallel.

The sequence has type: "[(int)]" (a sequence of sequences of integers). Given nested sequences and the rule that any function can be applied in parallel over the elements of a sequence, NESL necessarily supplies the ability to apply a parallel function multiple times in parallel; we call this ability nested parallelism. For example, we could apply the parallel sequence function \(\text{sum}\) over a nested sequence:

\[
\{\text{sum}(v) : v \in (\{2, 1\}, \{7, 3, 0\}, \{4\})\};
\]

\[
\Rightarrow [3, 10, 4] : [\text{int}]
\]

In this expression there is parallelism both within each \(\text{sum}\), since the sequence function has a parallel implementation, and across the three instances of \(\text{sum}\), since the apply-to-each construct is defined such that all instances can run in parallel.

NESL supplies a handful of functions for moving between levels of nesting. These include \(\text{flatten}\), which takes a nested sequence and flattens it by one level. For example,

\[
\text{flatten}((\{2, 1\}, \{7, 3, 0\}, \{4\}));
\]

\[
\Rightarrow [2, 1, 7, 3, 0, 4] : [\text{int}]
\]

Another useful function is \(\text{bottop}\) (for bottom and top), which takes a sequence of values and creates a nested sequence of length 2 with all the elements from the bottom half of the input sequence in the first element and elements from the top half in the second element (if the length of the sequence is odd, the bottom part gets the extra element). For example,

\[
\text{bottop}("\text{nested parallelism}"");
\]

\[
\Rightarrow ["\text{nested pa}", "\text{parallelism}"] : [[\text{char}]]
\]
Table 1 lists several examples of routines that could take advantage of nested parallelism. Nested parallelism also appears in most divide-and-conquer algorithms. A divide-and-conquer algorithm breaks the original data into smaller parts, applies the same algorithm on the subparts, and then merges the results. If the subproblems can be executed in parallel, as is often the case, the application of the subparts involves nested parallelism. Table 2 lists several examples.

As an example, consider how the function sum might be implemented,

```
function my_sum(a) =
    if (#a == 1) then a[0]
    else
        let r = {my_sum(v) : v in bottop(a)};
        in r[0] + r[1];
```

This code tests if the length of the input is one, and returns the single element if it is. If the length is not one, it uses bottop to split the sequence in two parts, and then applies itself recursively to each part in parallel. When the parallel calls return, the two results are extracted and added. The code effectively creates a tree of parallel calls which has depth  \( \log n \), where \( n \) is the length of \( a \), and executes a total of \( n - 1 \) calls to +.

As another more involved example, consider a parallel variation of quicksort [4] (see Figure 3). When applied to a sequence \( s \), this version splits the values into three subsets (the elements lesser, equal and greater than the pivot) and calls itself recursively on the lesser and greater subsets. To execute the two recursive calls, the lesser and greater sequences are concatenated into a nested sequence and \( qsort \) is applied over the two elements of the

---

3To simulate the built-in sum, it would be necessary to add code to return the identity (0) for empty sequences.
function qsort(a) =
if (#a < 2) then a
else
  let pivot = a[#a/2];
  lesser = {e in a | e < pivot};
  equal = {e in a | e == pivot};
  greater = {e in a | e > pivot};
  result = {qsort(v): v in [lesser, greater]}
in result[0] ++ equal ++ result[1];

Figure 3: An implementation of quicksort.

Figure 4: The quicksort algorithm. Just using parallelism within each block yields a parallel running time at least as great as the number of blocks (O(n)). Just using parallelism from running the blocks in parallel yields a parallel running time at least as great as the largest block (O(n)). By using both forms of parallelism the parallel running time can be reduced to the depth of the tree (expected O(lg n)).

nested sequences in parallel. The final line extracts the two results of the recursive calls and appends them together with the equal elements in the correct order.

The recursive invocation of qsort generates a tree of calls that looks something like the tree shown in Figure 4. In this diagram, taking advantage of parallelism within each block as well as across the blocks is critical to getting a fast parallel algorithm. If we were only to take advantage of the parallelism within each quicksort to subselect the two sets (the parallelism within each block), we would do well near the root and badly near the leaves (there are n leaves which would be processed serially). Conversely, if we were only to take advantage of the parallelism available by running the invocations of quicksort in parallel (the parallelism between blocks but not within a block), we would do well at the leaves and badly at the root (it would take n time to process the root). In both cases the parallel time complexity is O(n) rather than the ideal O(lg² n) we can get using both forms (this
is discussed in Section 1.5).

1.3 Pairs

As well as sequences, NESL supports the notion of pairs. A pair is a structure with two elements, each of which can be of any type. Pairs are often used to build simple structures or to return multiple values from a function. The binary comma operator is used to create pairs. For example:

```plaintext
9.8,"foo";
⇒ (9.8,"foo") : (float, [char])

2,3;
⇒ (2,3) : (int, int)
```

The comma operator is right associative (e.g. (2,3,4,5) is equivalent to (2, (3, (4,5))))

All other binary operators in NESL are left associative. The precedence of the comma operator is lower than any other binary operator, so it is usually necessary to put pairs within parentheses.

Pattern matching inside of a let construct can be used to deconstruct structures of pairs. For example:

```plaintext
let (a,b,c) = (2*4,5-2,4)
in a+b*c;
⇒ 20 : int
```

In this example, a is bound to 8, b is bound to 3, and c is bound to 4.

Nested pairs differ from sequences in several important ways. Most importantly, there is no way to operate over the elements of a nested pair in parallel. A second important difference is that the elements of a pair need not be of the same type, while elements of a sequence must always be of the same type.

1.4 Types

NESL is a strongly typed polymorphic language with a type inference system. Its type system is similar to functional languages such as ML, but since it is first-order (functions cannot be passed as data), function types only appear at the top level. Type variables of polymorphic functions can therefore range over all the data-types. As well as parametric polymorphism NESL also allows a form of overloading similar to what is supplied by the Haskell Language [21].

The type of a polymorphic function in NESL is specified by using type-variables, which are declared in a type-context. For example, the type of the `permute` function is:

```plaintext
[A], [int] -> [A] :: A in any
```
This specifies that for \( A \) bound to any type, \( \text{permute} \) maps a sequence of type \([A]\) and a sequence of type \([\text{int}]\) into another sequence of type \([A]\). The variable \( A \) is a type-variable, and the specification \( A \text{ in any} \) is the context. A context can have multiple type bindings separated by semicolons. For example, the pair function described in the last section has type:

\[
A, B \rightarrow (A,B) :: A \text{ in any}; B \text{ in any}
\]

User defined functions can also be polymorphic. For example we could define

\[
\text{function append3}(s1,s2,s3) = s1 ++ s2 ++ s3;
\]

\[
\Rightarrow \text{append3}(s1,s2,s3) : [A], [A], [A] \rightarrow [A] :: A \text{ in any}
\]

The type inference system will always determine the most general type possible.

In addition to parametric polymorphism, NESL supports a form of overloading by including the notion of \textit{type-classes}. A type-class is a set of types along with an associated set of functions. The functions of a class can only be applied to the types from that class. For example the base types, \text{int} and \text{float} are both members of the type class \text{number}, and numerical functions such as \(+\) and \(*\) are defined to work on all numbers. The type of a function overloaded in this way, is specified by limiting the context of a type-variable to a specific type-class. For example, the type of \(+\) is:

\[
A, A \rightarrow A :: A \text{ in number}
\]

The context "\( A \text{ in number} \)" specifies that \( A \) can be bound to any member of the type-class \text{number}. The fully polymorphic specification \text{any} can be thought of as type-class that contains all data types as members. The type-classes are organized into the hierarchy as shown in Figure 5. Functions such as \( = \) and \( < \) are defined on ordinals, functions such as \(+\) and \(*\) are defined on numbers, and functions such as \( \&\& \) or and \( \text{not} \) are defined on logicals.

User-defined functions can also be overloaded. For example:

\[
\text{function double}(a) = a + a;
\]

\[
\Rightarrow \text{double}(a) : A \rightarrow A :: A \text{ in number}
\]
It is also possible to restrict the type of a user-defined function by explicitly typing it. For example,

```plaintext
function double(a) : int -> int = a + a;
⇒ double(a) : int -> int
```

limits the type of `square` to `int -> int`. The `:` specifies that the next form is a type-specifier (see Appendix A for the full syntax of the function construct and type specifiers).

In certain situations the type inference system cannot determine the type even though there is one. For example the function:

```plaintext
function badfunc(a,b) = a or (a + b);
```

will not type properly because `or` is defined on the type-class logical and `+` is defined on the type-class number. As it so happens, `int` is both a logical and an integer, but the NESL inference system does not know how to take intersections of type-classes. In this situation it is necessary to specify the type:

```plaintext
function goodfunc(a,b) : int, int -> int = a or (a + b);
⇒ goodfunc(a,b) : int, int -> int
```

This situation comes up quite rarely.

Specifying the type using `:` serves as good documentation for a function even when the inference system can determine the type. The notion of type-classes in NESL is similar to the type-classes used in the Haskell language [21], but, unlike Haskell, NESL currently does not permit the user to add new type classes.\(^3\)

### 1.5 Deriving Complexity

There are two complexities associated with all computations in NESL.

1. **Work complexity**: this represents the total work done by the computation, that is to say, the amount of time that the computation would take if executed on a serial random access machine. The work complexity for most of the sequence functions is simply the size of one of its arguments. A complete list is given in Table 1. The size of an object is defined recursively: the size of a scalar value is 1, and the size of a sequence is the sum of the sizes of its elements plus 1.

2. **Step complexity**: this represents the parallel depth of the computation, that is to say, the amount of time the computation would take on a machine with an unbounded number of processors. The step complexity of all the sequence functions supplied by NESL is constant.

The work and step complexities are based on the vector random access machine (VRAM) model [5], a strictly data-parallel abstraction of the parallel random access machine (PRAM) model [16]. Since the complexities are meant for determining asymptotic complexity, these

\(^3\)It is likely that future versions of NESL will allow such extensions.
complexities do not include constant factors. All the NESL functions, however, can be executed in a small number of machine instructions per element.

The complexities are combined using simple combining rules. Expressions are combined in the standard way—for both the work complexity and the step complexity, the complexity of an expression is the sum of the complexities of the arguments plus the complexity of the call itself. For example, the complexities of the computation:

\[
\text{sum} (\text{dist}(7,n)) \ast \#a
\]

can be calculated as:

<table>
<thead>
<tr>
<th></th>
<th>Work</th>
<th>Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>dist</td>
<td>n</td>
<td>1</td>
</tr>
<tr>
<td>sum</td>
<td>n</td>
<td>1</td>
</tr>
<tr>
<td># (length)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>*</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>(O(n))</td>
<td>(O(1))</td>
</tr>
</tbody>
</table>

The apply-to-each construct is combined in the following way. The work complexity is the sum of the work complexity of the instantiations, and the step complexity is the maximum over the step complexities of the instantiations. If we denote the work required by an expression \(\exp\) applied to some data \(a\) as \(W(\exp(a))\), and the steps required as \(S(\exp(a))\), these combining rules can be written as

\[
W(\{e1(a) : a \text{ in } e2(b)\}) = W(e2(b)) + \text{sum}(\{W(e1(a)) : a \text{ in } e2(b)\})
\]

\[
S(\{e1(a) : a \text{ in } e2(b)\}) = S(e2(b)) + \text{max_val}(\{S(e1(a)) : a \text{ in } e2(b)\})
\]

where \text{sum} and \text{max_val} just take the sum and maximum of a sequence, respectively.

As an example, the complexities of the computation:

\[
\{[0:i] : i \text{ in } [0:n]\}
\]

can be calculated as:

<table>
<thead>
<tr>
<th></th>
<th>Work</th>
<th>Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0:n]</td>
<td>n</td>
<td>1</td>
</tr>
<tr>
<td>Parallel Calls</td>
<td>(\sum_{i=0}^{n} i)</td>
<td>(\max_{i=0}^{n} 1)</td>
</tr>
<tr>
<td>Total</td>
<td>(O(n^2))</td>
<td>(O(1))</td>
</tr>
</tbody>
</table>

Once the work \((W)\) and step \((S)\) complexities have been calculated in this way, the formula

\[
T = O(\frac{W}{P} + S \lg P)
\]

14
places an upper bound on the asymptotic running time of an algorithm on the CRCW
PRAM model \((P)\) is the number of processors). This formula can be derived from Brent's
scheduling principle [10] as shown in [34, 5, 23]. The \(\lg P\) term shows up because of the
cost of allocating tasks to processors, and the cost of implementing the sum and scan
operations. On the scan-PRAM [4], where it is assumed that the scan operations are no
more expensive than references to the shared-memory (they both require \(O(\lg P)\) on a
machine with bounded degree circuits), then the equation is:

\[
T = O(W/P + S)
\]  

(4)

In the mapping onto a PRAM, the only reason a concurrent-write capability is required is
for implementing the \(-\) (put) function, and the only reason a concurrent-read capability is
required is for implementing the \(-\) (get) function. Both of these functions allow repeated
indices ("collisions") and could therefore require concurrent access to a memory location.
If an algorithm does not use these functions, or guarantees that there are no collisions
when they are used, then the mapping can be implemented with a EREW PRAM. Out of
the algorithms in this paper, the primes algorithm (Section 2.2) requires concurrent writes,
and the string-searching algorithm (Section 2.1) requires concurrent reads. All the other
algorithms can use an EREW PRAM.

As an example of how the work and step complexities can be used, consider the
\texttt{kth smallest} algorithm described earlier (Figure 2). In this algorithm the work is the
same as the time required by the standard serial version (loops have been replaced by par-
allel calls), which has an expected time of \(O(n)\) [15]. It is also not hard to show that the
expected number of recursive calls is \(O(\lg n)\), since we expect to drop some fraction of the
elements on each recursive call [30]. Since each recursive call requires a constant number
of steps, we therefore have:

\[
W(n) = O(n) \quad S(n) = O(\lg n)
\]

Using Equation 3 this gives us an expected case running time on a PRAM of:

\[
T(n) = O(n/p + \lg n \lg p) = O(n/p + \lg^2 n) \quad \text{EREW PRAM}
\]

\[
= O(n/p + \lg n) \quad \text{scan-PRAM}
\]

One can similarly show for the quicksort algorithm given in Figure 3 that the work and step
complexities are \(W(n) = O(n \lg n)\) and \(S(n) = O(\lg n)\) [30], which give a EREW PRAM
running time of:

\[
T(n) = O(n \lg n/p + \lg^2 n) \quad \text{EREW PRAM}
\]

\[
= O(n \lg n/p + \lg n) \quad \text{scan-PRAM}
\]

In the remainder of this paper we will only derive the work and step complexities. The
reader can plug these into Equation 3 or Equation 4 to get the PRAM running times.
2 Examples

This section describes several examples of NESL programs. Before describing the examples we describe three common operations, get, put and integer sequences. The \(-\rightarrow\) binary operator (called get) is used to extract multiple elements from a sequence. Its left argument is the sequence to extract from, and the right argument is a sequence of integer indices which specify from which locations to extract elements. For example, the expression

\[
"an\ \text{example}\"\rightarrow[7,\ 0,\ 8,\ 4];
\]

\(\Rightarrow\ "pale" : [\text{char}]\)

extracts the p, a, 1 and e from locations 7, 0, 8 and 4, respectively. The \(-\leftarrow\) binary operator (called put) is used to insert multiple elements into a sequence. Its left argument is the sequence to insert into (the destination sequence) and its right argument is a sequence of integer-value pairs. For each element \((i,v)\) in the sequence of pairs, the value \(v\) is inserted into position \(i\) of the destination sequence. For example, the expression

\[
"an\ \text{example}\"\leftarrow[(4,\ 's),(2,\ 'd),(3,\ \text{space})];
\]

\(\Rightarrow\ "\text{and}\ \text{sample}" : [\text{char}]\)

inserts the s, d and space into the string "an example" at locations 4, 2 and 3, respectively (space is a constant that is bound to the space character).

Ranges of integers can be created using square brackets along with a colon. The notation

\([\text{start} : \text{end}]\)

creates a sequence of integers starting at \(\text{start}\) and ending one before \(\text{end}\). For example:

\([10:16];\)

\(\Rightarrow\ [10,\ 11,\ 12,\ 13,\ 14,\ 15] : [\text{int}]\)

An additional stride can be specified, as in \([\text{start}\ : \text{end}\ : \text{stride}]\), which returns every \(\text{stride}^{th}\) integer between \(\text{start}\) and \(\text{end}\). For example:

\([10:25:3];\)

\(\Rightarrow\ [10,\ 13,\ 16,\ 19,\ 22] : [\text{int}]\)

The integer \(\text{end}\) is never included in the sequence.

Using these operations, it is easy to define many of the other NESL functions. Figure 6 shows several examples.

2.1 String Searching

The first example is a function that finds all occurrences of a word in a string (a sequence of characters). The function \texttt{string\_search}(\(w, s\)) (see Figure 7) takes a search word \(w\) and a string \(s\), and returns the starting position of all substrings in \(s\) that match \(w\). For example,

\[
\text{string\_search}("\text{foo}\", "\text{fobarfoofoofoo}\");
\]

\(\Rightarrow\ [5, 12] : [\text{int}]\)
function subseq(a,start,end) = a->[start:end];

function take(a,n) = a->[0:n];

function drop(a,n) = a->[n:#a];

function rotate(a,n) = a->{aod(i-n,#a) : i in [n:n +#a]};

function even_elts(a) = a->[0:#a:2];

function odd_elts(a) = a->[1:#a:2];

function bottop(a) = [a->[0:#a/2],a->[#a/2:#a]];

Figure 6: Possible implementation for several of the NESL functions on sequences.

function next_character(candidates,w,s,i) =
  if (i == #w) then candidates
  else
    let letter = w[i];
    next_l = s->{c + i: c in candidates};
    candidates = {c in candidates; n in next_l | n == letter}
    in next_character(candidates, w, s, i+1);

function string_search(w, s) = next_character([0:#s - #w], w, s, 0);

Figure 7: Finding all occurrences of the word w in the string s.
The algorithm starts by considering all positions between 0 and \#s-\#w as candidates for a match (no candidate could be greater than this since it would have to match past the end of the string). The candidates are stored as pointers (indices) into s of the beginning of each match. The algorithm then progresses through the search string, using recursive calls to next_char, narrowing the set of candidate matches on each step.

Based on the current candidates, next_char narrows the set of candidates by only keeping the candidates that match on the next character of v. To do this, each candidate checks whether the i-th character in v matches the i-th position past the candidate index. All candidates that do match are packed and passed into the recursive call of next_char. The recursion completes when the algorithm reaches the end of v. The progression of cands in the "foo" example would be:

<table>
<thead>
<tr>
<th>i</th>
<th>cands</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>[0, 5, 8, 12]</td>
</tr>
<tr>
<td>1</td>
<td>[0, 5, 12]</td>
</tr>
<tr>
<td>2</td>
<td>[5, 12]</td>
</tr>
</tbody>
</table>

Let's consider the complexity of the algorithm. We assume \#w = m and \#s = n. The number of steps taken by the algorithm is some constant times the number of recursive calls, which is simply \(O(m)\). The work complexity of the algorithm is the sum over the calls of the number of candidates in each step. In practice, this is usually \(O(n)\), but in the worst case this can be the product of the two lengths \(O(nm)\) (the worst case can only happen if most of the characters in v are repeated). There are parallel string-searching algorithms that give better bounds on the parallel time (step complexity), and that bound the worst case work complexity to be linear in the length of the search string [11, 37], but these algorithms are somewhat more complicated.

### 2.2 Primes

Our second example finds all the primes less than n. The algorithm is based on the sieve of Eratosthenes. The basic idea of the sieve is to find all the primes less than \(\sqrt{n}\), and then use multiples of these primes to "sieve out" all the composite numbers less than n. Since all composite numbers less than n must have a divisor less than \(\sqrt{n}\), the only elements left unsieved will be the primes. There are many parallel versions of the prime sieve, and several naive versions require a total of \(O(n^{3/2})\) work and either \(O(n^{1/2})\) or \(O(n)\) parallel time. A well designed version should require no more work than the serial sieve \(O(n \log \log n)\), and polylogarithmic parallel time.

The version we use (see Figure 8) requires \(O(n \log \log n)\) work and \(O(\log \log n)\) steps. It works by first recursively finding all the primes up to \(\sqrt{n}\) (sqr.primes). Then, for each prime \(p\) in sqr.primes, the algorithm generates all the multiples of \(p\) up to \(n\) (sieves). This is done with the [\#:4] construct. The sequence sieves is therefore a nested sequence with each subsequence being the sieve for one of the primes in sqr.primes. The function flatten, is now used to flatten this nested sequence by one level, therefore returning a sequence containing all the sieves. For example,
function primes(n) =
if n == 2 then [2]
else
    let sqr_primes = primes(cei(sqrt(float(n))));
    sieves = [(2p:n:p): p in sqr_primes];
    flat_sieves = flatten(sieves);
    flags = dist(t,n) <- {(i,f): i in flat_sieves}
in drop({i in [0:n]: flag in flags| flag}, 2);

Figure 8: Finding all the primes less than n.

flazten(EC4, 6, 8, 10, 12, 14, 16, 18, [6, 9, 12, 15, 18]));
⇒ [4, 6, 8, 10, 12, 14, 16, 18, 6, 9, 12, 15, 18] : [int]

This sequence of sieves is used by the <- function to place a false flag in all positions that are a multiple of one of the sqr_primes. This will return a boolean sequence, flags, which contains a t in all places that were not knocked out by a sieve—these are the primes. However, we want primes to return the indices of the primes instead of flags. To generate these indices the algorithm creates a sequences of all indices between 0 and n ([0:n]) and uses subselection to remove the nonprimes. The function drop is then used to remove the first two elements (0 and 1), which are not considered primes but do not get explicitly sieved.

The functions [s:e:d], flatten, dist, <- and drop all require a constant number of steps. Since primes is called recursively on a problem of size sqrt(n) the total number of steps require by the algorithm can be written as the recurrence:

\[
S(n) = \begin{cases} 
O(1) & n = 1 \\
S(\sqrt{n}) + O(1) & n > 1 
\end{cases} = O(\log \log n)
\]

Almost all the work done by primes is done in the first call. In this first call, the work is proportional to the length of the sequence flat_sieves. Using the standard formula

\[
\sum_{p \leq x} 1/p = \log \log x + C + O(1/\log x)
\]

where p are the primes [18], the length of this sequence is:

\[
\sum_{p \leq \sqrt{n}} n/p = O(n \log \log \sqrt{n})
\]

= \(O(n \log \log n)
\)

therefore giving a work complexity of \(O(n \log \log n)\).
Figure 9: An example of the quickhull algorithm. Each sequence shows one step of the algorithm. Since $A$ and $P$ are the two $z$ extrema, the line $AP$ is the original split line. $J$ and $N$ are the farthest points in each subspace from $AP$ and are, therefore, used for the next level of splits. The values outside the brackets are hull points that have already been found.
function cross-product(o,line) =
let (xo,yo) = o;
((x1,y1),(x2,y2)) = line
in (x1-xo)*(y2-yo) - (y1-yo)*(x2-xo);

function hsplit(points,p1,p2) =
let cross = {cross-product(p,(p1,p2)) : p in points};
packed = {p in points; c in cross | plusp(c)}
in if (#packed < 2) then [p1] ++ packed
else
  let pm = points[max_index(cross)]
in flatten({hsplit(packed,p1,p2) : p1 in [p1,pm]; p2 in [pm,p2]});

function convex-hull(points) =
let x = {x : (x,y) in points};
  minx = points[min_index(x)];
  maxx = points[max_index(x)]
in hsplit(points,minx,maxx) ++ hsplit(points,maxx,minx);

Figure 10: Code for Quickhull. Each point is represented as a pair. Pattern matching is used to extract the x and y coordinates of each pair.

2.3 Planar Convex-Hull

Our next example solves the planar convex hull problem: given n points in the plane, find which of these points lie on the perimeter of the smallest convex region that contains all points. The planar convex hull problem has many applications ranging from computer graphics [17] to statistics [20]. The algorithm we use to solve the problem is a parallel version [8] of the quickhull algorithm [29]. The quickhull algorithm was given its name because of its similarity to the quicksort algorithm. As with quicksort, the algorithm picks a "pivot" element, splits the data based on the pivot, and is recursively applied to each of the split sets. Also, as with quicksort, the pivot element is not guaranteed to split the data into equally sized sets, and in the worst case the algorithm will require $O(n^2)$ work.

Figure 9 shows an example of the quickhull algorithm, and Figure 10 shows the code. The algorithm is based on the recursive routine hsplit. This function takes a set of points in the plane ((x,y) coordinates) and two points p1 and p2 that are known to lie on the convex hull, and returns all the points that lie on the hull clockwise from p1 to p2, inclusive of p1, but not of p2. In Figure 9, given all the points [A, B, C, ..., P], p1 = A and p2 = P, hsplit would return the sequence [A, B, J, O]. In hsplit, the order of p1 and p2 matters, since if we switch A and P, hsplit would return the hull along the other direction [P, W, C].

The hsplit function works by first removing all the elements that cannot be on the hull since they lie below the line between p1 and p2. This is done by removing elements whose cross product with the line between p1 and p2 are negative. In the case p1 = A and p2 =
P, the points [B, D, F, G, H, J, K, M, O] would remain and be placed in the sequence packed. The algorithm now finds the point furthest from the line p1-p2. This point pm must be on the hull since as a line at infinity parallel to p1-p2 moves toward p1-p2, it must first hit pm. The point pm (J in the running example) is found by taking the point with the maximum cross-product. Once pm is found, hsplit calls itself twice recursively using the points (p1, pm) and (pm, p2) ((A, J) and (J, P) in the example). When the recursive calls return, hsplit flattens the result (this effectively appends the two subhulls).

The overall convex-hull algorithm works by finding the points with minimum and maximum x coordinates (these points must be on the hull) and then using hsplit to find the upper and lower hull. Each recursive call has a step complexity of \(O(1)\) and a work complexity of \(O(n)\). However, since many points might be deleted on each step, the work complexity could be significantly less. For \(m\) hull points, the algorithm runs in \(O(\lg m)\) steps for well-distributed hull points, and has a worst case running time of \(O(m)\) steps.

3 Language Definition

This section defines NESL. It is not meant as a formal semantics but, along the full definition of the syntax in Appendix A and description of all the built-in functions in Appendix B, it should serve as an adequate description of the language. NESL is a strict first-order strongly-typed language with the following data types:

- four primitive atomic data types: booleans (bool), integers (int), characters (char), and floats (float);
- the primitive sequence type;
- the primitive pair type;
- and user definable compound datatypes;

and the following operations:

- a set of predefined functions on the primitive types;
- three primitive constructs: a conditional construct if, a binding construct let, and the apply-to-each construct;
- and a function constructor, function, for defining new functions.

This section covers each of these topics.

3.1 Data

3.1.1 Atomic Data Types

There are four primitive atomic data types: booleans, integers, characters and floats.

The boolean type bool can have one of two values t or f. The standard logical operations (eg. not, and, or, xor, nor, nand) are predefined. The operations and, or, xor, nor, nand all use infix notation. For example:
The integer type int is the set of (positive and negative) integers that can be represented in the fixed precision of a machine-sized word. The exact precision is machine dependent, but will always be at least 32-bits. The standard functions on integers (+, -, *, /, ==, >, <, negate, ...) are predefined, and use infix notation (see Appendix A for the precedence rules). For example:

```
3 * -11;
⇒ -33 : int
```

Overflow will return unpredictable results.

The character type char is the set of ASCII characters. The characters have a fixed order and all the comparison operations (eg. ==, <, >, ...) can be used. Characters are written by placing a ' in front of the character. For example:

```
'a < 'd;
⇒ t : bool
```

The global variables space, newline and tab are bound to the appropriate characters.

The type float is used to specify floating-point numbers. The exact representation of these numbers is machine specific, but NESL tries to use 64-bit IEEE when possible. Floats support most of the same functions as integers, and also have several additional functions (eg. round, truncate, sqrt, log,...). Floats must be written by placing a decimal point in them so that they can be distinguished from integers.

```
1.2 * 3.0;
⇒ 3.6 : float
```

23
There is no implicit coercion between scalar types. To add 2 and 3.0, for example, it is
necessary to coerce one of them: e.g.

\[
\text{float(2) + 3.0;}
\]

\[
\Rightarrow 5.0 : \text{float}
\]

A complete list of the functions available on scalar types can be found in Appendix B.1.

3.1.2 Sequences \(\square\)

A sequence is an ordered set of values. A sequence can contain any type, including other
sequences, but each element in a sequence must be of the same type (sequences are homo-
genous). The type of a sequence whose elements are of type \(\alpha\), is specified as \([\alpha]\). For
examples:

\[
[6, 2, 4, 5];
\]

\[
\Rightarrow [6, 2, 4, 5] : [\text{int}]
\]

\[
[[2, 1, 7, 3], [6, 2], [22, 9]];\]

\[
\Rightarrow [[2, 1, 7, 3], [6, 2], [22, 9]] : [[\text{int}]]
\]

Sequences of characters can be written between double quotes,

"a string";

\[
\Rightarrow "a string" : [\text{char}]
\]

but can also be written as a sequence of characters:

['a, Space, 's, 't, 'r, 'i, 'n, 'g];

\[
\Rightarrow "a string" : [\text{char}]
\]

Empty sequences must be explicitly typed since the type cannot be determined from
the elements. The type of an empty sequences is specified by using empty square braces
followed by the type of the elements. For example,

\[
\square \text{int;}
\]

\[
\Rightarrow \square : [\text{int}]
\]

\[
\square (\text{int, bool});
\]

\[
\Rightarrow \square : [(\text{int, bool})]
\]

Appendix B.2 describes the functions that operate on sequences.
3.1.3 Record Types (datatype)

Record types with a fixed number of slots can be defined with the `datatype` construct. For example,

```
datatype complex(float, float);
⇒ complex(a1, a2) : float, float → complex
```

defines a record with two slots both which must contain a floating-point number. Defining a record also defines a corresponding function that is used to construct the record. For example,

```
complex(7.1, 11.9);
⇒ complex(7.1, 11.9) : complex
```

creates a `complex` record with 7.1 and 11.9 as its two values. The type of the record is specified as `complex()`.

Elements of a record can be accessed using pattern matching in the `let` construct. For example,

```
let complex(real, imaginary) = a
in real;
```

will remove the real part of the variable `a` (assuming it is kept in the first slot). More details on pattern matching are given in the next section.

As with functions, records can be parameterized based on type-variables. For example, `complex` could have been defined as:

```
datatype complex(alpha, alpha) :: alpha in number;
⇒ complex(a1, a2) : alpha, alpha → complex(alpha) :: alpha in number
```

This specifies that for `alpha` bound to any type in the type-class `number` (either `int` or `float`), both slots must be of type `alpha`. This will allow either,

```
complex(7.1, 11.9);
⇒ complex(7.1, 11.9) : complex(float)
complex(7, 11);
⇒ complex(7, 11) : complex(int)
```

but will not allow `complex(7, 'a)` or `complex(2, 2.2)`. The type of a record is specified by the record name followed by the binding of all its type-variables. In this case, the binding of the type-variable is either `int` or `float`.
3.2 Functions and Constructs

3.2.1 Conditionals (if)

The only primitive conditional in NESL is the if construct. The syntax is:

\[
\text{IF } exp \text{ THEN } exp \text{ ELSE } exp
\]

If the first expression is true, then the second expression is evaluated and its result is returned, otherwise the third expression is evaluated and its result is returned. The first expression must be of type bool, and the other two expressions must be of identical types.

For example:

\[
\text{if } (t \text{ and } f) \text{ then } 3 + 4 \text{ else } (6 - 2) * 7
\]

is a valid expression, but

\[
\text{if } (t \text{ and } f) \text{ then } 3 \text{ else } 2.6
\]

is not, since the two branches return different types.

3.2.2 Binding Local Variables (let)

Local variables can be bound with the let construct. The syntax is:

\[
\text{LET } expbinds \text{ IN } exp
\]

\[
\begin{align*}
\text{expbinds} &::= \text{ expbind } [; \text{ expbinds}] \quad \text{variable bindings} \\
\text{expbind} &::= \text{ pattern } = \text{ exp} \quad \text{variable binding} \\
\text{pattern} &::= \text{ ident} \quad \text{variable} \\
&\quad \text{ident(pattern)} \quad \text{datatype pattern} \\
&\quad ( \text{ pattern } ) \quad \text{pair pattern}
\end{align*}
\]

The semicolon separates bindings (the square brackets indicate an optional term of the syntax). Each pattern is either a variable name or a pattern based on a record name. Each expbind binds the variables in the pattern on the left of the = to the result of the expression on the right. For example:

\[
\text{let a = 7; (b, c) = (1, 2) in a*(b + c)};
\]

\[
\Rightarrow 21 : \text{int}
\]

Here a is bound to 7, then the pattern (b, c) is matched with the result of the expression on the right so that b is bound to 1 and c is bound to 2. Patterns can be nested, and the patterns are matched recursively.

The variables in each expbind can be used in the expressions (exp) of any later expbind (the bindings are done serially). For example, in the expression
let a = 7;
    b = a + 4
in a * b;
⇒ 77 : int

the variable a is bound to the value 7 and then the variable b is bound to the value of a
plus 4, which is 11. When these are multiplied in the body, the result is 77.

3.2.3 The Apply-to-Each Construct ({})

The apply-to-each construct is used to apply any function over the elements of a sequence.
It has the following syntax:

\[
\{[\text{exp} : \text{rbinds} [:\text{exp}]]
\]

\[
\text{rbinds} ::= \text{rbind [:; \text{rbinds}}
\]

\[
\text{rbind} ::= \text{pattern IN exp} \quad \text{full binding}
\]

\[
\text{ident} \quad \text{shorthand binding}
\]

An apply-to-each construct consists of three parts: the expression before the colon, which
we will call the body, the bindings that follow the body, and the expression that follows the
\[
\]
which we will call the sieve. Both the body and the sieve are optional: they could both
be left out, as in

\[
\{a \text{ in } [1, 2, 3];
\Rightarrow [1, 2, 3] : [\text{int}]
\]

The \text{rbinds} can contain multiple bindings which are separated by semicolons. We first
consider the case in which there is a single binding. A binding can either consist of a pattern
followed by the keyword IN and an expression (full binding), or consist of a variable name
(shorthand binding). In a full binding the expression is evaluated (it must evaluate to a
sequence) and the variables in the pattern are bound in turn to each element of the sequence.
The body and sieve are applied for each of these bindings. For example:

\[
\{a + 2: a \text{ in } [1, 2, 3];
\Rightarrow [3, 4, 5] : [\text{int}]
\]

\[
\{a + b: (a,b) \text{ in } [(1,2), (3,4), (5,6)];
\Rightarrow [3, 7, 11] : [\text{int}]
\]

In a shorthand binding, the variable must be a sequence, and the body and sieve are applied
to each element of the sequence with the variable name bound to the element. For example:

\[
\text{let } a = [1, 2, 3]
\text{in } \{a + 2: a; \\
\Rightarrow [3, 4, 5] : [\text{int}]
\]
In the case of multiple `rbinds`, each of the sequences (either the result of the expression in a full binding or the value of the variable in a shorthand binding) must be of equal length. The bindings are interleaved so that the body is evaluated with bindings made for elements at the same index of each sequence. For example:

\[
\{a + b: a \text{ in } [1, 2, 3]; b \text{ in } [1, 4, 9]\};
\]

\[
\Rightarrow [2, 6, 12] : [\text{int}]
\]

\[
\{\text{dist}(b, a): a \text{ in } [1, 2, 3]; b \text{ in } [1, 4, 9]\};
\]

\[
\Rightarrow [[[1], [4, 4], [9, 9, 9]] : [[[\text{int}]]]
\]

An apply-to-each with a body and two bindings,

\[
\{\text{body: pattern1 in exp1; pattern2 in exp2 | sieve}\}
\]

is equivalent to the single binding construct

\[
\{\text{body: (pattern1,pattern2) in zip(exp1,exp2) | sieve}\}
\]

where `zip`, as defined in the list of functions, elementwise zips together the two sequences it is given as arguments.

If there is no body in an apply-to-each construct, then the results of the first binding is returned. For example:

\[
\{a \text{ in } [1, 2, 3]; b \text{ in } [1, 4, 9]\};
\]

\[
\Rightarrow [1, 2, 3] : [\text{int}]
\]

\[
\{a \text{ in } [1, 2, 3]; b \text{ in } [2, 4, 9] | b = 2*a\};
\]

\[
\Rightarrow [1, 2] : [\text{int}]
\]

\[
\{b \text{ in } [2, 4, 9]; a \text{ in } [1, 2, 3] | b = 2*a\};
\]

\[
\Rightarrow [2, 4] : [\text{int}]
\]

If there is a body and a sieve, the body and sieve are both evaluated for all bindings, and then the subselection is applied. An apply-to-each with a sieve of the form:

\[
\{\text{body : bindings | sieve}\}
\]

is equivalent to the construct

\[
\text{pack}(([\text{body, sieve}) : \text{bindings}])
\]

where `pack`, as defined in the list of functions, takes a sequence of type `[((\text{alpha},\text{bool})]]` and returns a sequence which contains the first element of each pair if the second element is true. The order of remaining elements is maintained.
3.2.4 Defining New Functions (function)

Functions can be defined at top-level using the function construct. The syntax is:

\[
\text{FUNCTION ident pattern [: funtype]} = \text{exp} ;
\]

A function has one argument, but the argument can be any pattern. The body of a function (the exp at the end) can only refer to variables bound in the pattern, or variables declared at top-level. Any function referred to in the body can only refer to functions previously defined or to the function itself (at present there is no way to define mutually recursive functions). As with all functional languages, defining a function with the same name as a previous function only hides the previous function from future use: all references to a function before the new definition will refer to the original definition.

3.2.5 Top-Level Bindings (=)

You can bind a variable at top-level using the = operator. The syntax is:

\[
\text{ident} = \text{exp} ;
\]

For example, \(a = 211\); will bind the variable \(a\) to the value 211. The variable can now either be referenced at top level, or can be referenced inside of any function. For example, the definition

\[
\text{function foo(c) = c + a;}
\]

would define a function that adds 211 to its input. Such top-level binding is mostly useful for saving temporary results at top-level, and for defining constants. The variable \(\pi\) is bound at top level to the value of \(\pi\).

Acknowledgments

I would like to thank Marco Zagha, Jay Sipelstein, Margaret Reid-Miller, Bob Harper, Jonathan Hardwick, John Greiner, Tim Freeman, and Siddhartha Chatterjee for many helpful comments on this manual. Siddhartha Chatterjee, Jonathan Hardwick, Jay Sipelstein, and Marco Zagha did all the work getting the intermediate languages VCODE and CVL running so that NESL can actually run on parallel machines.

References


A The NESL Grammar

This appendix defines the grammar of NESL. The grammatical conventions are:

- The brackets [] enclose optional phrases, the symbol * means repeat the previous expression any number of times, and the symbol + means repeat the previous expression any number of times, but at least once.

- All symbols in typewriter font are literal tokens, all symbols in boldface are tokens with the lexical definitions given below, and all symbols in italics are variables (nonterminals) of the grammar.

- All uppercase letters can either be upper or lower case. NESL is case insensitive.

Toplevel

toplevel ::= FUNCTION ident pattern [: funtype] = exp ; function definition
DATATYPE ident typeexp [: : typebinds] ; datatype definition
ident = exp ; variable binding
typeexp ; expression

Types

funtype ::= typeexp -> typeexp [: : typebinds] function type
typebinds ::= typebind [, typebinds] binding type variables
typebind ::= ident IN typeclass binding a type variable
typeexp ::= basetype base type
ident type variable
ident([typelist]) compound datatype
typeexp , typeexp pair type
typeexp ', typeexp ']' sequence type
( typeexp )
typelist ::= typeexp [, typelist] type list
typeclass ::= NUMBER | ORDINAL | LOGICAL | ANY the type classes
basetype ::= INT | BOOL | FLOAT | CHAR the base types
Expressions

\[
\begin{align*}
\text{exp} &::= \text{const} \\
&\quad \text{ident} \\
&\quad \text{IF exp THEN exp ELSE exp} \\
&\quad \text{LET expbinds IN exp} \\
&\quad \{[exp : ] \text{rbinds [1 exp]}\} \\
&\quad \text{ident(exp)} \\
&\quad \text{exp binop exp} \\
&\quad \text{unaryop exp} \\
&\quad \text{sequence} \\
&\quad \text{exp ['' exp '']} \\
&\quad (\text{exp}) \\
\end{align*}
\]
constant
variable
conditional
local bindings
apply-to-each
function application
binary operator
unary operator
sequence
sequence extraction
parenthesized expression

\[
\begin{align*}
\text{expbinds} &::= \text{pattern = exp [; expbinds]} \\
\text{pattern} &::= \text{ident} \\
&\quad \text{ident(pattern)} \\
&\quad \text{pattern, pattern} \\
&\quad (\text{pattern}) \\
\text{rbinds} &::= \text{rbind [; rbinds]} \\
\text{rbind} &::= \text{pattern IN exp} \\
&\quad \text{ident} \\
\text{sequence} &::= \text{'' explist '}
\text{'' exp : exp [; exp] '}
\text{empty sequence} \\
\text{intconst} \\
\text{floatconst} \\
\text{boolean (T or F)} \\
\text{character string} \\
\text{precedence 1} \\
\text{precedence 2} \\
\text{precedence 3} \\
\text{precedence 4} \\
\text{precedence 5} \\
\text{precedence 6} \\
\text{precedence 7} \\
\text{precedence 8}
\end{align*}
\]
Lexical Definitions

The following defines regular expressions for the lexical classes of tokens. The grammatical conventions are:

- All uppercase letters can either be upper or lower case. NESL is case insensitive.

- The brackets ( ) enclose an expression. The brackets [ ] enclose a character set, any one of which must match. The expression 0-9 within square brackets means all digits and the expression A-Z means all letters. The symbol ~ as the first character within square brackets means a compliment character set (all characters excepting the following ones).

- The symbol * means the previous expression can be repeated as many times as needed, the symbol + means the previous expression can be repeated as many times as needed but at least once, and the symbol ? means the previous expression can be matched either once or not at all.

\[
\begin{align*}
\text{intconst} & ::= \ [-+]? [0-9]^* \\
\text{floatconst} & ::= \ [-+]? [0-9]^* \cdot [0-9]^* ([eE] [-+]? [0-9]^*)^? \\
\text{ident} & ::= \ [_A-Z0-9]^+ \\
\text{boolconst} & ::= \ [TF] \\
\text{stringconst} & ::= \ "[\"][\"]*"
\end{align*}
\]


**B List of Functions**

This section lists the functions available in NESL. Each function is listed in the following way:

\[
\text{function interface} \quad \{\text{source-types} \rightarrow \text{result-type : type-bindings}\}
\]

Definition of function.

The hierarchy of the type classes is shown in Figure 5.

**B.1 Scalar Functions**

Logical Functions

All the logical functions work on either integers or booleans. In the case of integers, they work bitwise over the bit representation of the integer.

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>not(a)</code></td>
<td>Returns the logical inverse of the argument. For integers, this is the ones complement.</td>
<td>(a \rightarrow a : a \text{ in logical})</td>
</tr>
<tr>
<td><code>a or b</code></td>
<td>Returns the inclusive or of the two arguments.</td>
<td>(a,a \rightarrow a : a \text{ in logical})</td>
</tr>
<tr>
<td><code>a and b</code></td>
<td>Returns the logical and of the two arguments.</td>
<td>(a,a \rightarrow a : a \text{ in logical})</td>
</tr>
<tr>
<td><code>a xor b</code></td>
<td>Returns the exclusive or of the two arguments.</td>
<td>(a,a \rightarrow a : a \text{ in logical})</td>
</tr>
<tr>
<td><code>a nor b</code></td>
<td>Returns the inverse of the inclusive or of the two arguments.</td>
<td>(a,a \rightarrow a : a \text{ in logical})</td>
</tr>
<tr>
<td><code>a nand b</code></td>
<td>Returns the inverse of the and of the two arguments.</td>
<td>(a,a \rightarrow a : a \text{ in logical})</td>
</tr>
</tbody>
</table>

Comparison Functions

All comparison functions work on integers, floats and characters.

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>a == b</code></td>
<td>Returns (t) if the two arguments are equal.</td>
<td>(a,a \rightarrow \text{bool} : a \text{ in ordinal})</td>
</tr>
<tr>
<td><code>a /= b</code></td>
<td>Returns (t) if the two arguments are not equal.</td>
<td>(a,a \rightarrow \text{bool} : a \text{ in ordinal})</td>
</tr>
</tbody>
</table>
\(a < b\) \\(\{a,a \rightarrow bool : a \text{ in ordinal}\}\)
Returns \(t\) if the first argument is strictly less than the second argument.

\(a > b\) \\(\{a,a \rightarrow bool : a \text{ in ordinal}\}\)
Returns \(t\) if the first argument is strictly greater than the second argument.

\(a \leq b\) \\(\{a,a \rightarrow bool : a \text{ in ordinal}\}\)
Returns \(t\) if the first argument is less than or equal to the second argument.

\(a \geq b\) \\(\{a,a \rightarrow bool : a \text{ in ordinal}\}\)
Returns \(t\) if the first argument is greater or equal to the second argument.

**Predicates**

plusp(v) \\(\{a \rightarrow bool : a \text{ in number}\}\)
Returns \(t\) if \(v\) is strictly greater than 0.

minusp(v) \\(\{a \rightarrow bool : a \text{ in number}\}\)
Returns \(t\) if \(v\) is strictly less than 0.

zerop(v) \\(\{a \rightarrow bool : a \text{ in number}\}\)
Returns \(t\) if \(v\) is equal to 0.

oddp(v) \(\{int \rightarrow bool\}\)
Returns \(t\) if \(v\) is odd (not divisible by two).

evenp(v) \(\{int \rightarrow bool\}\)
Returns \(t\) if \(v\) is even (divisible by two).

**Arithmetic Functions**

\(a + b\) \\(\{a,a \rightarrow a : a \text{ in number}\}\)
Returns the sum of the two arguments.

\(a - b\) \\(\{a,a \rightarrow a : a \text{ in number}\}\)
Subtracts the second argument from the first.

\(-v\) \\(\{a \rightarrow a : a \text{ in number}\}\)
Negates a number.

\(\text{abs}(x)\) \\(\{a \rightarrow a : a \text{ in number}\}\)
Returns the absolute value of the argument.
\textbf{diff}(x, y) \quad \{a,a \rightarrow a : a \text{ in number}\}

Returns the absolute value of the difference of the two arguments.

\textbf{max}(a, b) \quad \{a,a \rightarrow a : a \text{ in ordinal}\}

Returns the argument that is greatest (closest to positive infinity).

\textbf{min}(a, b) \quad \{a,a \rightarrow a : a \text{ in ordinal}\}

Returns the argument that is least (closest to negative infinity).

\textbf{v * d} \quad \{a,a \rightarrow a : a \text{ in number}\}

Returns the product of the two arguments.

\textbf{v / d} \quad \{a,a \rightarrow a : a \text{ in number}\}

Returns \(v\) divided by \(d\). If the arguments are integers, the result is truncated towards 0.

\textbf{rem}(v, d) \quad \{int,int \rightarrow int\}

Returns the remainder after dividing \(v\) by \(d\).

\textbf{lshift}(a, b) \quad \{int,int \rightarrow int\}

Returns the first argument logically shifted to the left by the integer contained in the second argument. Shifting will fill with 0-bits.

\textbf{rshift}(a, b) \quad \{int,int \rightarrow int\}

Returns the first argument logically shifted to the right by the integer contained in the second argument. Shifting will fill with 0-bits or the sign bit, depending on the implementation.

\textbf{sqrt}(v) \quad \{float \rightarrow float\}

Returns the square root of the argument. The argument must be nonnegative.

\textbf{isqrt}(v) \quad \{int \rightarrow int\}

Returns the greatest integer less than or equal to the exact square root of the integer argument. The argument must be nonnegative.

\textbf{ln}(v) \quad \{float \rightarrow float\}

Returns the natural log of the argument.

\textbf{log}(v, b) \quad \{float,float \rightarrow float\}

Returns the logarithm of \(v\) in the base \(b\).

\textbf{exp}(v) \quad \{float \rightarrow float\}

Returns \(e\) raised to the power \(v\).
\[ \text{exp}(v, p) \]  
Returns \( v \) raised to the power \( p \).

\[ \sin(v) \]  
Returns the sine of \( v \), where \( v \) is in radians.

\[ \cos(v) \]  
Returns the cosine of \( v \), where \( v \) is in radians.

\[ \tan(v) \]  
Returns the tangent of \( v \), where \( v \) is in radians.

\[ \text{asin}(v) \]  
Returns the arc sine of \( v \). The result is in radians.

\[ \text{acos}(v) \]  
Returns the arc cosine of \( v \). The result is in radians.

\[ \text{atan}(v) \]  
Returns the arc tangent of \( v \). The result is in radians.

\[ \sinh(v) \]  
Returns the hyperbolic sine of \( v \) \( \frac{(e^v - e^{-v})}{2} \).

\[ \cosh(v) \]  
Returns the hyperbolic cosine of \( v \) \( \frac{(e^v + e^{-v})}{2} \).

\[ \tanh(v) \]  
Returns the hyperbolic tangent of \( v \) \( \frac{(e^v - e^{-v})}{(e^v + e^{-v})} \).

**Conversion Functions**

\[ \text{btoi}(a) \]  
Converts the boolean values \( t \) and \( f \) into 1 and 0, respectively.

\[ \text{code.char}(a) \]  
Converts an integer to a character. The integer must be the code for a valid character.

\[ \text{char.code}(a) \]  
Converts a character to its integer code.

\[ \text{float}(v) \]  
\( \{ \text{int} \rightarrow \text{float} \} \)
Converts an integer to a floating-point number.

\texttt{ceil(v)} \quad \{\text{float} \rightarrow \text{int}\}

Converts a floating-point number to an integer by truncating toward positive infinity.

\texttt{floor(v)} \quad \{\text{float} \rightarrow \text{int}\}

Converts a floating-point number to an integer by truncating toward negative infinity.

\texttt{trunc(v)} \quad \{\text{float} \rightarrow \text{int}\}

Converts a floating-point number to an integer by truncating toward zero.

\texttt{round(v)} \quad \{\text{float} \rightarrow \text{int}\}

Converts a floating-point number to an integer by rounding to the nearest integer; if the number is exactly halfway between two integers, then it is implementation specific to which integer it is rounded.

\section*{Other Scalar Functions}

\texttt{rand(v)} \quad \{a \rightarrow a : a \text{ in number}\}

For a positive value \(v\), \texttt{rand} returns a random value in the range \([0..v)\).

\section*{B.2 Sequence Functions}

\subsection*{Simple Sequence Functions}

\texttt{#v} \quad \{[a] \rightarrow \text{int} : a \text{ in any}\}

Returns the length of a sequence.

\texttt{dist(a, 1)} \quad \{a, \text{int} \rightarrow [a] : a \text{ in any}\}

Generates a sequence of length 1 with the value \(a\) in each element. For example:

\begin{align*}
a &= a_0 \\
1 &= 5 \\
\text{dist}(a, 1) &= [a_0, a_0, a_0, a_0, a_0]
\end{align*}

\texttt{a[i]} \quad \{[a, \text{int} \rightarrow a : a \text{ in any}\}

Extracts the element specified by index \(i\) from the sequence \(a\). Indices are zero-based.

\texttt{rep(d, v, i)} \quad \{[a], \text{int} \rightarrow [a] : a \text{ in any}\}

Replaces the \(i\)th value in the sequence \(d\) with the value \(v\). For example:
d = \{a_0, a_1, a_2, a_3, a_4\}
v = b_0
i = 3
rep(d, v, i) = \{a_0, a_1, a_2, b_0, a_4\}

zip(a, b) = \{(b, a) : a \in \text{any}; b \in \text{any}\}
Zips two sequences of equal length together into a single sequence of pairs.

**Scans and Reduces**

plus.scan(a) = \{[a] \rightarrow [a] : a \in \text{number}\}
Given a sequence of numbers, plus.scan returns to each position of a new equal-length sequence, the sum of all previous positions in the source. For example:

\[
a = [1, 3, 5, 7, 9, 11, 13, 15]\\
\text{plus.scan}(a) = [0, 1, 4, 9, 16, 25, 36, 49]
\]

max.scan(a) = \{[a] \rightarrow [a] : a \in \text{ordinal}\}
Given a sequence of ordinals, max.scan returns to each position of a new equal-length sequence, the maximum of all previous positions in the source. For example:

\[
a = [3, 2, 1, 6, 5, 4, 8]\\
\text{max.scan}(a) = [-\infty, 3, 3, 3, 6, 6, 6]
\]

min.scan(a) = \{[a] \rightarrow [a] : a \in \text{ordinal}\}
Given a sequence of ordinals, min.scan returns to each position of a new equal-length sequence, the minimum of all previous positions in the source.

or.scan(a) = \{[\text{bool}] \rightarrow [\text{bool}]\}
A scan using logical-or on a sequence of booleans.

and.scan(a) = \{[\text{bool}] \rightarrow [\text{bool}]\}
A scan using logical-and on a sequence of booleans.

\([s:e:d]\) = \{\text{int, int, int} \rightarrow [\text{int}]\}
Returns a set of indices starting at s, increasing by d, and finishing before e. For example:

\[
s = 4\\
d = 3\\
e = 15\\
\text{[s:e:d]} = [4, 7, 10, 13]
\]
sum(v) \{ [a] \rightarrow a : a \text{ in number} \}

Given a sequence of numbers, \( \text{sum} \) returns their sum. For example:

\[ v = [7, 2, 9, 11, 3] \]
\[ \text{sum}(v) = 32 \]

max_val(v) \{ [a] \rightarrow a : a \text{ in ordinal} \}

Given a sequence of ordinals, \( \text{max_val} \) returns their maximum.

min_val(v) \{ [a] \rightarrow a : a \text{ in ordinal} \}

See max_val.

any(v) \{ [\text{bool}] \rightarrow \text{bool} \}

Given a sequence of booleans, \( \text{any} \) returns \( \text{t} \) iff any of them are \( \text{t} \).

all(v) \{ [\text{bool}] \rightarrow \text{bool} \}

Given a sequence of booleans, \( \text{all} \) returns \( \text{t} \) iff all of them are \( \text{t} \).

count(v) \{ [\text{bool}] \rightarrow \text{int} \}

Counts the number of \( \text{t} \) flags in a boolean sequence. For example:

\[ v = [\text{T}, \text{F}, \text{T}, \text{T}, \text{F}, \text{T}, \text{F}, \text{T}] \]
\[ \text{count}(v) = 5 \]

max_index(v) \{ [a] \rightarrow \text{int} : a \text{ in ordinal} \}

Given a sequence of ordinals, \( \text{max_index} \) returns the index of the maximum value. If several values are equal, it returns the leftmost index. For example:

\[ v = [2, 11, 4, 7, 14, 6, 9, 14] \]
\[ \text{max_index}(v) = 4 \]

min_index(v) \{ [a] \rightarrow \text{int} : a \text{ in ordinal} \}

Given a sequence of ordinals, \( \text{min_index} \) returns the index of the minimum value. If several values are equal, it returns the leftmost index.

Sequence Reordering Functions

values -> indices \{ [a], [\text{int}] \rightarrow [a] : a \text{ in any} \}

Given a sequence of \( \text{values} \) on the left and a sequence of \( \text{indices} \) on the right, which can be of different lengths, \( \rightarrow \) returns a sequence which is the same length as the \( \text{indices} \) sequence and the same type as the \( \text{values} \) sequence. For each position in the \( \text{indices} \) sequence, it extracts the value at that index of the \( \text{values} \) sequence. For example:
values = [a₀, a₁, a₂, a₃, a₄, a₅, a₆, a₇]
indices = [3, 5, 2, 6]
values -> indices = [a₃, a₅, a₂, a₆]

permute(v, i)  
{[a],[int] → [a] : a in any}

Given a sequence v and a sequence of indices i, which must be of the same length, permute permutes the values to the given indices. The permutation must be one-to-one.

d <- ivpairs  
{[a],[((int,a))] → [a] : a in any}

This operator, called put, is used to insert multiple elements into a sequence. Its left argument is the sequence to insert into (the destination sequence) and its right argument is a sequence of integer-value pairs. For each element (i, v) in the sequence of integer-value pairs, the value v is inserted into position i of the destination sequence.

rotate(a, i)  
{[a],[int] → [a] : a in any}

Given a sequence and an integer, rotate rotates the sequence around by i positions to the right. If the integer is negative, then the sequence is rotated to the left. For example:

\[a = [a₀, a₁, a₂, a₃, a₄, a₅, a₆, a₇]\]
\[i = 3\]
\[rotate(a, i) = [a₅, a₆, a₇, a₀, a₁, a₂, a₃, a₄]\]

reverse(a)  
{[a] → [a] : a in any}

Reverses the order of the elements in a sequence.

Simple Sequence Manipulation

pack(v)  
{[(a,bool)] → [a] : a in any}

Given a sequence of (value, flag) pairs, pack packs all the values with a t in their corresponding flag into consecutive elements, deleting elements with an f.

v₁ ++ v₂  
{[a],[a] → [a] : a in any}

Given two sequences, ++ appends them. For example:

\[v₁ = [a₀, a₁, a₂]\]
\[v₂ = [b₀, b₁]\]
\[v₁ ++ v₂ = [a₀, a₁, a₂, b₀, b₁]\]

subseq(v, start, end)  
{[a],[int,int] → [a] : a in any}

Given a sequence, subseq returns the subsequence starting at position start and ending one before position end. For example:
\[
v = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7] \\
\text{start} = 2 \\
\text{end} = 6 \\
\text{subseq}(v, \text{start}, \text{end}) = [a_2, a_3, a_4, a_5]
\]

\[
\text{drop}(v, n) = \{[a], \text{int} \rightarrow [a] : a \text{ in any}\}
\]

Given a sequence, drop drops the first \(n\) items from the sequence. For example:

\[
v = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7] \\
\text{n} = 3 \\
\text{drop}(v, n) = [a_3, a_4, a_5, a_6, a_7]
\]

\[
\text{take}(v, n) = \{[a], \text{int} \rightarrow [a] : a \text{ in any}\}
\]

Given a sequence, take takes the first \(n\) items from the sequence. For example:

\[
v = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7] \\
\text{n} = 3 \\
\text{take}(v, n) = [a_0, a_1, a_2]
\]

\[
\text{odd.slts}(v) = \{[a] \rightarrow [a] : a \text{ in any}\}
\]

Returns the odd indexed elements of a sequence.

\[
\text{even.slts}(v) = \{[a] \rightarrow [a] : a \text{ in any}\}
\]

Returns the even indexed elements of a sequence.

\[
\text{interleave}(a, b) = \{[a], [a] \rightarrow [a] : a \text{ in any}\}
\]

Interleaves the elements of two sequences. The sequences must be of the same length. For example:

\[
a = [a_0, a_1, a_2, a_3] \\
b = [b_0, b_1, b_2, b_3] \\
\text{interleave}(a, b) = [a_0, b_0, a_1, b_1, a_2, b_2, a_3, b_3]
\]

**Nesting Sequences**

The two functions partition and flatten are the primitives for moving between levels of nesting. All other functions for moving between levels of nesting can be built out of these. The functions split and bottop are often useful for divide-and-conquer routines.

\[
\text{partition}(v, \text{counts}) = \{[a], \text{int} \rightarrow [[a]] : a \text{ in any}\}
\]

Given a sequence of values and another sequence of counts, partition returns a nested sequence with each subsequence being of a length specified by the counts. The sum of the counts must equal the length of the sequence of values. For example:
\(v = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7]\)
\(\text{counts} = [4, 1, 3]\)
\(\text{partition}(v, \text{counts}) = [[a_0, a_1, a_2, a_3], [a_4], [a_5, a_6, a_7]]\)

\(\text{flatten}(v)\) \(\{[[a]] \rightarrow [a] : a \in \text{any}\}\)
Given a nested sequence of values, \(\text{flatten}\) flattens the sequence. For example:
\(v = [[a_0, a_1, a_2], [a_3, a_4], [a_5, a_6, a_7]]\)
\(\text{flatten}(v) = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7]\)

\(\text{split}(v, \text{flags})\) \(\{[a],[\text{bool}] \rightarrow [[a]] : a \in \text{any}\}\)
Given a sequence of values \(a\) and a boolean sequence of flags, \(\text{split}\) creates a nested sequence of length 2 with all the elements with an \(f\) in their flag in the first element and elements with a \(t\) in their flag in the second element. For example:
\(v = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7]\)
\(\text{flags} = [T, F, T, F, T, T, T]\)
\(\text{split}(v, \text{flags}) = [[a_1, a_3, a_4], [a_0, a_2, a_5, a_6, a_7]]\)

\(\text{bottop}(v)\) \(\{[a] \rightarrow [[a]] : a \in \text{any}\}\)
Given a sequence of values \(\text{values}\), \(\text{bottop}\) creates a nested sequence of length 2 with all the elements from the bottom half of the sequence in the first element and elements from the top half of the sequence in the second element. For example:
\(v = [a_0, a_1, a_2, a_3, a_4, a_5, a_6]\)
\(\text{bottop}(v) = [[a_0, a_1, a_2, a_3], [a_4, a_5, a_6]]\)

\(\text{head\_rest}(\text{values})\) \(\{[a] \rightarrow a,[a] : a \in \text{any}\}\)
Given a sequence of values \(\text{values}\) of length > 0, \(\text{head\_rest}\) returns a pair containing the first element of the sequence, and the remaining elements of the sequence.

\(\text{rest\_tail}(\text{values})\) \(\{[a] \rightarrow [a],a : a \in \text{any}\}\)
Given a sequence of values \(\text{values}\) of length > 0, \(\text{rest\_tail}\) returns a pair containing all but the last element of the sequence, and the last element of the sequence.

**Other Sequence Functions**

These are more complex sequence functions. The step complexities of these functions are not \(O(1)\).

\(\text{sort}(a)\) \(\{[\text{int}] \rightarrow [\text{int}]\}\)
Sorts the input sequence. The sort is stable—equal elements will maintain their relative order.

**rank(a)**  
{[int] → [int]}

Returns the rank of each element of the sequence a. The rank of an element is the position it would appear in if the sequence were sorted. A sort of a sequence a can be implemented as `permute(a, rank(a))`. The rank is stable.

**collect(key-value_pairs)**  
{[(a,b)] → [(a, b)]: a in any; b in any}

Takes a sequence of (key, value) pairs, and collects each set of values that have the same key together into a sequence. The function returns a sequence of (key, value-sequence) pairs. Each key will only appear once in the result and the value-sequence corresponding to the key will contain all the values that had that key in the input.

**kth_smallest(s, k)**  
{[a], int → a : a in ordinal}

Returns the kth smallest element of a sequence s (k is 0 based). It uses the quick-select algorithm and therefore has expected work complexity of $O(n)$ and an expected step complexity of $O(\log n)$.

**search_for_subseqs(subseq, sequence)**  
{[a], [a] → [int]: a in any}

Returns indices of all start positions in sequence where the string specified by subseq appears.

**remove_duplicates(s)**  
{[a] → [a]: a in any}

Removes duplicates from a sequence. Elements are considered duplicates if eql on them returns T.

**union(a, b)**  
{[a], [a] → [a]: a in any}

Given two sequences each which has no duplicates, union will return the union of the elements in the sequences.

**intersection(a, b)**  
{[a], [a] → [a]: a in any}

Given two sequences each which has no duplicates, intersection will return the intersection of the elements in the sequences.

**name(a)**  
{[a] → [int]: a in any}

This function assigns an integer label to each unique value of the sequence a. Equal values will always be assigned the same label and different values will always be assigned different labels. All the labels will be in the range [0..#a) and will correspond to the position in a of one of the elements with the same value. The function `remove_duplicates(a)` could be implemented as `{s in a; i in [0:#a]; r in name(a) | r == i}`.
B.3 Functions on Any Type

**eql(a, b)** \( \{a, a \rightarrow \text{bool} : a \in \text{any}\} \)

Given two objects of the same type, `eql` will return `t` if they are equal and `f` otherwise. Two sequences are equal if they are the same length and their elements are elementwise equal. Two records are equal if their fields are equal.

**0v** \( \{a \rightarrow [\text{char}] : a \in \text{any}\} \)

Given any printable object `v`, `0` converts it into its printable representation as a character string.

**hash(a, 1)** \( \{a, \text{int} \rightarrow \text{int} : a \in \text{any}\} \)

Hashes the argument `a` and returns an integer in the range [0..1).

**select(flag, v1, v2)** \( \{\text{bool}, a, a \rightarrow a : a \in \text{any}\} \)

Returns the second argument if the flag is `T` and the third argument if the flag is `F`. This differs from an if form in that both arguments are evaluated.

B.4 Functions with Side Effects

The functions in this section are not purely functional. Unless otherwise noted, none of them can be called in parallel—they cannot be called within an apply-to-each construct. The routines in this section are not part of the core language, they are meant for debugging, I/O, timing and display. Because these functions are new it is reasonably likely that the interface of some of these functions will change in future versions. The user should check the most recent documentation.

Input and Output Routines

Of the functions listed in this section, only `print_char`, `print_string`, `printf_char`, `printf_string`, and `print_debug` can be called in parallel.

**print_char(v)** \( \{\text{char} \rightarrow \text{bool}\} \)

Prints a character to standard output

**print_string(v)** \( \{[\text{char}] \rightarrow \text{bool}\} \)

Prints a character string to standard output

**print_debug(str, v)** \( \{[\text{char}], a \rightarrow a : a \in \text{any}\} \)

Prints the character string `str` followed by the string representation of the object `v`, and then a newline to standard output. This function can be useful when debugging.

**write_object_to_file(object, filename)** \( \{a, [\text{char}] \rightarrow a : a \in \text{any}\} \)

Writes an object to a file. The first argument is the object and the second argument is
a filename. For example `write.object.to.file([2,3,1,0],"/tmp/foo")` would write a vector of integers to the file `/tmp/foo`. The data is stored in an internal format and can only be read back using `read.object.from.file`.

```plaintext
write.string.to.file(a, filename)  {[char],[char] → [char]}
Prints a character string to the file named filename.
```

```plaintext
read.object.from.file(object_type, filename)  {a,[char] → a : a in any}
Reads an object from a file. The first argument is an object of the same type as the object to be read, and the second argument is a filename from which the object will be read. For example, the call `read.object.from.file(C,"/tmp/foo")` would read an integer from the file `/tmp/foo`, and `read.object.from.file([I"int","/tmp/bar")` would read a vector of integers from the file `/tmp/foo`. The object needs to have been stored using the function `write.object.to.file`.
```

```plaintext
read.int.seq.from.file(filename)  {[char] → [int]}
Reads a sequence of integers from the file named filename. The file must start with a left parenthesis, contain the integers separated by either whitespaces, newlines or tabs, and end with a right parenthesis. For example:

```
( 22 33 11
10 14
12 11 )
```
represents the sequence [22, 33, 11, 10, 14, 12, 11].
```

```plaintext
read.float.seq.from.file(filename)  {[char] → [float]}
Reads a sequence of floats from the file named filename. The file must start with a left parenthesis, contain the floats separated by either whitespaces, newlines or tabs, and end with a right parenthesis. The file may contain integers (no .); these will be coerced to floats.
```

```plaintext
open.out.file(filename)  {[char] → file.pointer}
Opens a file for writing and returns a file-pointer to that file. File pointers cannot be returned to top-level. They must be used within a single top-level call.
```

```plaintext
close.file(filep)  {file.pointer → bool}
Closes a file given a file_pointer. It returns T if the close was successful and F otherwise.
```

```plaintext
printf.char(a, filep)  {char,file.pointer → bool}
Prints a character to the file pointed to by the file pointer filep.
```

```plaintext
printf.string(a, filep)  {[char],file.pointer → bool}
Prints a character string to the file pointed to by the file pointer filep.
```

48
Plotting Functions

The functions in this section can be used for plotting data on an Xwindow display. Currently no color plotting is supported.

\[ \text{make-window}((x_0, y_0), \text{width}, \text{height}), \text{bbox}, \text{display}) \]

\[ \{(\text{int}, \text{int}, \text{int}, \text{boundingbox}, \text{char}) \rightarrow \text{window}\} \]

Creates a window on the display specified by display. Its upper left hand corner will be at position \((x_0, y_0)\) on the screen and will have a size as specified by width and height. The bbox argument specifies the bounding box for the data to be plotted in the window. The bounding box is a structure that specifies the virtual coordinates of the window. It can be created with the function bounding-box. Note that windows get automatically closed when you return to top-level. This means that you cannot return a window to top-level and then use it—you must create it and use it within a single top-level call.

\[ \text{bounding-box}((\text{points})) \]

\[ \{(a, b)\} \rightarrow \text{boundingbox} : a \text{ in number}; b \text{ in number}\} \]

Creates a bounding-box to be used by make-window. Given a sequence of points, this box is determined by the maximum and minimum x and y values.

\[ \text{draw-points}((\text{points}, \text{window})) \]

\[ \{(b, a), \text{window} \rightarrow \text{int} : a \text{ in number}; b \text{ in number}\} \]

Draws a sequence of points into the window specified by window. The window must have been created by make-window.

\[ \text{draw-lines}((\text{points}, \text{window})) \]

\[ \{(b, a), \text{window} \rightarrow \text{int} : a \text{ in number}; b \text{ in number}\} \]

Draws a sequence of lines between the points in the points argument into the window specified by window.

\[ \text{draw-segments}((\text{segs}, \text{window})) \]

\[ \{((c, d), a, b), \text{window} \rightarrow \text{int} : a \text{ in number}; b \text{ in number}; c \text{ in number}; d \text{ in number}\} \]

Draws a sequence of line segments into the window specified by window. Each line-segment is specified as a pair of points.

\[ \text{close-window}((\text{window})) \]

\[ \{\text{window} \rightarrow \text{int}\} \]

Closes a window created with make-window. After executing this command, the window will not accept any more of the draw commands, and will go away if you mouse on it.

Other Side Effecting Functions

\[ \text{time}(a) \]

\[ \{a \rightarrow a, \text{float} : a \text{ in any}\} \]

The expression \(\text{TIme}(exp)\) returns a pair whose first element is the value of the expression \(exp\) and whose second element is the time in seconds taken to execute the expression \(exp\). This function cannot be called in parallel (within an apply-to-each).
C Implementation Notes

This section describes some hints for writing efficient code in NESL. Most of these hints are based on the current implementation: tradeoffs are likely to change in future implementations. The section also points out some deficiencies with the current implementation.

The Read-Eval-Print Loop

Here we outline how the interactive environment of NESL works. This should give the user a feeling in some cases for where time is going. When the user types an expression at top-level, the following steps take place:

1. The expression gets compiled into the intermediate language VCODE.

2. All code in the expression's call tree that has not been previously compiled gets compiled into VCODE. When you define a function in NESL it only gets partially compiled immediately—the compilation completes the first time it is called. Because of this delayed compilation, calling a function can take longer the first time it is used.

3. The VCODE for the expression and all functions in its call tree get written to a file. This file can actually be inspected by the user, if so desired (see the user’s manual).

4. The environment starts up a subprocess that executes the VCODE interpreter on the VCODE file. The VCODE interpreter is a stand-alone executable program.

5. When the interpreter completes the computation, it writes the output to a new file and exits.

6. When the interpreter has finished writing the output, the NESL environment reads the output file, interprets the data and prints the result.

When executing on a remote machine, the only step that differs is Step 4. Instead of executing the interpreter locally, the environment executes the appropriate version of the interpreter remotely (using rsh). If the remote machine is on a shared file system, such as AFS, then no files need to be explicitly copied. If it is not on a shared file system, then the VCODE file gets copied by the system to the remote machine before execution and the results get copied back when the interpreter completes.

Using Large Data Sets

In the current implementation of NESL users need to be somewhat careful about returning large data-sets to the top-level interpreter, or of passing large data sets in as an argument at top-level (we consider a data-set large if it contains more that 10,000 or so elements). Such passing can be quite slow since it require writing the data out to a file and then reading it back in. To avoid this bottleneck, we suggest using one of the NESL I/O functions to read and write the data (e.g. read_object_from_file, read_int_seq_from_file and write_object_to_file).
For example, if a user had an application `solve` that required a large sequence of pairs as input, and returned another large sequence of pairs as output, the best way to write this would be:

```plaintext
function solve_from_file(infilename, outfilename)
let in_data = read_object_from_file([], (int,int), infilename)
result = solve(in_data);
tmp = write_object_to_file(result, outfilename)
in take(result, 100)
```

Note that `solve_from_file` function only returns the first 100 elements of the result. This makes it possible to make sure the result looks reasonable without returning the whole thing, which would be slow. Instead the whole result gets written to a file.

The Truth about Complexity

Equations 1 and 2 in Section 1.5 specified how the work and step complexities could be combined in an apply-to-each. In the current implementation there are a couple caveats to this, which the user should be aware of. The first concerns work complexity. In the following discussion we will consider a variable constant with regards to an apply-to-each if the variable is free (not bound) in the body of the apply-to-each and is not defined in bindings of the apply-to-each. For example, in

```
{foo(a,b*c): b in s}
```

the variables `a` and `c` are free with regards to the apply-to-each, while `b` is not. We will refer to these variables as free-vars. In the current implementation all free-vars need to be copied across the instances of an apply-to-each. This copying requires time, and the equation for combining work complexity that includes this cost is:

\[
W(\{e1(a): a in e2(b)\}) = W(e2(b)) + \sum_{c \in \text{free-vars}} (\text{Length}(e2(b)) \times \text{Size}(c))
\]

where the last term is has been added to Equation 1 (`Length(e2(b))` refers to the length of the sequence returned by `e2(b)`, and `Size(c)` refers to the size of each free-var). If a free-var is large, this copy could be the dominant cost of an apply-to-each. Here are some examples of such cases:

<table>
<thead>
<tr>
<th>Expression</th>
<th>Work Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>{#a + i : i in a}</code></td>
<td>$(#a)^2$</td>
</tr>
<tr>
<td><code>{a[i] : i in b}</code></td>
<td>$#a \times #b$</td>
</tr>
</tbody>
</table>

In both cases the work is a factor of $#a$ greater than we might expect since the sequence `a` needs to be copied over the instances. As well as requiring extra work these copies require significant extra memory and can be a memory bottleneck in a program. Both the above examples can easily be rewritten to reduce the work and memory:
The user should be conscious of these costs and rewrite such expressions.

A second problem with the current implementation is that Equation 2 (the combining rule for step complexity) only holds if the body of the apply-to-each is contained. The definition of contained code is code where only one branch of a conditional has a non constant step complexity. For example, the following function is not contained because both branches of the inner if have $S(n) > O(1)$:

```plaintext
def function power(a, n) =
    if (n == 0) then 1
    else
        if evenp(n)
            then square(power(a, n/2))
        else a * square(power(a, n/2))
```

This can be fixed by calculating power(a, n/2) outside the conditional:

```plaintext
def function power(a, n) =
    if (n == 0) then 1
    else
        let pow = power(a, n/2)
        in if evenp(n)
            then square(pow)
        else a * square(pow)
```

In future implementations of NESL it is likely that this restriction will be removed.
Index

*, 38
+, 37
++, 43
-, 37
->, 42
/, 38
/=, 36
<, 37
<=, 43
<=, 37
=, 29, 33
==, 36
>, 37
>=, 37
@, 47
#, 40
abs, 37
acos, 39
all, 42
and, 36
and_scan, 41
any, 42
Apply-to-each, 27
asin, 39
Assignment, 29
atan, 39
Booleans (bool), 22
bottop, 17, 45
bounding_box, 49
btoi, 39
ceil, 40
char_code, 39
Characters (char), 23
collect, 46
Complexity, 13
Convex hull, 21
cos, 39
cosh, 39
count, 42
datatype, 25, 33
diff, 38
dist, 40
draw_lines, 49
draw_points, 49
draw_segments, 49
drop, 17, 44
e1t, 40
eql, 47
even_elts, 17, 44
evenp, 37
exp, 38
expn, 39
flatten, 45
float (function), 40
Floats (type), 23
floor, 40
function, 29, 33
Global variables, 29
hash, 47
head_rest, 45
if, 34
Integers (int), 23
interleave, 44
intersection, 46
iseq, 41
isqrt, 38
kth_smallest, 7, 46
last, 26, 34
ln, 38
Local bindings, 26
Logical, 12
log, 38
lshift, 38
make_window, 49
max, 38
max_index, 42
max_scan, 41
max_val, 42
min, 38
min_index, 42
min_scan, 41
min_val, 42
minusp, 37
name, 46
nand, 36
Nested parallelism, 7
nor, 36
not, 36
Number, 12
odd_elts, 17, 44
oddp, 37
open_out_file, 48
time, 49
Toplevel, 33
truce, 40
Type classes, 12
Types, 11, 33
union, 46
Work complexity, 13
write_object_to_file, 47
write_string_to_file, 48
xor, 36
zerop, 37
zip, 41

ord, 36
or_scan, 41
Ordinal, 12
pack, 43
partition, 44
Patterns, 34
permute, 43
plus_scan, 41
plusp, 37
Precedence, 34
Prime numbers, 18
primes, 19
print_char, 47
print_debug, 47
print_string, 47
printf_char, 48
printf_string, 48
Quickhull, 21
Quicksort, 10
rand, 40
rank, 46
read_float_seq_from_file, 48
read_int_seq_from_file, 48
read_object_from_file, 48
Records, 25
remove_duplicates, 46
rem, 38
rep, 40
rest_tail, 45
reverse, 43
rotate, 17, 43
round, 40
rshift, 38
search_for_subseqs, 46
select, 47
Sequences, 24, 34
sin, 39
sinh, 39
sort, 46
split, 45
sqrt, 38
Step complexity, 13
String searching, 16
Strings, 24
subseq, 17, 43
sum, 42
take, 17, 44
tan, 39
tanh, 39
This report describes NESL, a strongly-typed, applicative, data-parallel language. NESL is intended to be used as a portable interface for programming a variety of parallel and vector supercomputers, and as a basis for teaching parallel algorithms. Parallelism is supplied through a simple set of data-parallel constructs based on sequences (ordered sets), including a mechanism for applying any function over the elements of a sequence in parallel and a rich set of parallel functions that manipulate sequences.

NESL fully supports nested sequences and nested parallelism — the ability to take a parallel function and apply it over multiple instances in parallel. Nested parallelism is important for implementing algorithms with complex and dynamically changing data structures, such as required in many graph and sparse matrix algorithms. NESL also provides a mechanism for calculating the asymptotic running time for a program on various parallel machine models, including the parallel random access machine (PRAM). This is useful for estimating running times of algorithms on actual machines and, when teaching algorithms, for supplying a close correspondence between the code and the theoretical complexity.

This report defines NESL and describes several examples of algorithms coded in the language. The examples include algorithms for median finding, sorting, string searching, finding prime numbers, and finding a planar convex hull. NESL currently compiles to an intermediate language called Vcode, which runs on the Cray Y-MP, Connection Machine CM-2, and Encore Multimax. For many algorithms, the current implementation gives performance close to optimized machine-specific code for these machines.

Note: This report is an updated version of CMU-CS-92-103, which described version 2.4 of the language. The most significant changes in version 2.6 are that it supports polymorphic types, has an ML-like syntax instead of a lisp-like syntax, and includes support for I/O.

Keywords: DATA-PARALLEL, PARALLEL ALGORITHMS, SUPERCOMPUTERS, NESTED PARALLELISM, PRAM MODEL, PARALLEL PROGRAMMING LANGUAGES, COLLECTION ORIENTED

(54 pages)
Carnegie Mellon University does not discriminate and Carnegie Mellon University is required not to discriminate in admission, employment or administration of its programs on the basis of race, color, national origin, sex, handicap, or sex of handicap, in violation of Title VI of the Civil Rights Act of 1964. It also complies with the Educational Amendments of 1972 and Section 504 of the Rehabilitation Act of 1973, including honoring state or local laws or executive orders.

In addition, Carnegie Mellon University does not discriminate on the basis of sex or gender in the administration of its programs on the basis of race, color, national origin, sex, handicap, or sex of handicap, in violation of Title VI of the Civil Rights Act of 1964. It also complies with the Educational Amendments of 1972 and Section 504 of the Rehabilitation Act of 1973, including honoring state or local laws or executive orders.

Inquiries concerning application of these statements should be directed to the Director of Admissions and Financial Aid, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890, telephone (412) 268-2000.