*Macromol. Symp.* ***182,*** *195–207 (2002)*

# A New Arithmetic for Linear Free Radical Copolymerization

*Jon A. Debling* [1] *and Fouad Teymour* [2*]

[1]Johnson Polymer,8310 16th Street, PO Box 902,Sturtevant, WI,
53177-0902, USA
[2]Department of Chemical and Environmental Engineering Illinois Institute of Technology, Chicago, IL, 60616, USA

**Summary:** As a consequence of their method of production, polymer chains are polydisperse in size, composition and sequence distribution. In this work we present a new method of uniquely identifying these “polymer isomers“ termed “Digital Encoding of Polymeric Chains“. The method involves replacing distinguishable features of the chain such as monomer units, branches, etc. with a number. This unique sequence of numbers provides a digital code, which, depending on the base of the arithmetic used (binary, ternary) can be translated into a unique decimal equivalent number. We have applied this technique to the case of binary copolymerization in a CSTR at steady state and show how the sequence spectra of the chain populations are conveniently obtained. Furthermore, the technique shows that rich information about the copolymerization kinetics, reactivity ratios and termination mode can be obtained from analysis of the short chains of the distribution. The implications for this in parameter estimation and controlled polymerization are discussed in this paper.

## Introduction

Polymeric materials have had a significant impact on the development of most modern technological advancements. The diversity of material properties that polymers offer is the main reason behind their successful usage in a variety of applications. For example, the application properties of homopolymers can be manipulated by changes in their molecular weight distribution and/or branching characteristics. Copolymers on the other hand offer added flexibility as one can alter not only their composition, but also the monomer sequencing details on their chains. In all cases, the successful usage of polymeric materials in applications has been limited to those properties that derive from their distributed nature. As opposed to small molecule materials, which are usually utilized in high-purity forms or in well-defined blends and composites, a polymer is typically produced as a collection of molecules (chains) that differ in molecular weight, molecular formula, composition and structure. For some applications, this nature is advantageous since polymer properties can thus portray a multitude of characteristics. For example, the presence of a low molecular weight fraction in polyolefins facilitates

© WILEY-VCH Verlag GmbH, 69469 Weinheim, 2002
CCC 1022-1360/00/$ 17.50+.50/0

their processing, but this fraction alone does not have the correct performance characteristics required of these polymers. On the other hand, one could argue that some applications benefit only from the properties contributed by a sub-grouping of chains present in a polymer sample, while other chains could either be totally uninvolved or actually lessening the magnitude of display of the desired property. There must then exist another end for the spectrum at which pure polymeric components have a tremendous untapped potential to provide previously unobserved novel properties beyond the capabilities of presently known polymeric materials. It is the achievement of this level of specificity in producing or purifying polymers that will thrust the polymer industry into a new age of discovery not unlike the rapid developments at the turn of the last century. Before this can be achieved however, new methods have to be developed to enable the description of polymeric chains at the specific level of distinction of what we have termed “polymeric isomers”. The present article introduces a new modeling technique developed for this purpose.

Since polymerization is statistical in nature, the polymer chains in a copolymer sample are polydisperse in size, composition and sequence distribution. Thus, any two chains of the same length and composition are not necessarily identical molecules, despite the fact that they possess the same molecular formula. The difference stems from the order of monomer sequences along them, and in this regard is similar to the difference between small isomeric molecules, like 2,3-dimethylpentane and 2,4-dimethylpentane. Methods that acknowledge the exact sequence order on the chain will allow for the distinction of "polymeric isomers" which, like the two forms of dimethylpentane, could have vastly different properties. At any given chain length, all polymer isomers are probable but with different probabilities that depend on kinetics and process conditions. Since each isomeric form may have distinct properties, manipulation of the rate of their formation can lead to advantageous polymer property control. Current research in controlled polymerization using RAFT, ATRP and metallic catalysts and recent interest in nanocomposites, along with future advances in polymerization will heighten the interest in producing polymers with precisely defined architectures. This concept is widely accepted in biotechnology, as it is well understood that the precise sequencing and unique structures of DNA and protein molecules are responsible for their very specific roles in the functioning of living organisms.

Traditionally, the analysis of chain sequence information in copolymerization systems has been limited to the calculation of the sequence length distribution (SLD), for sequences of different lengths, using probabilistic arguments (c.f. [1] for a complete list of references). This is usually compared to SLD measurements by NMR techniques. While this combined methodology is very useful for the purpose of differentiation between copolymers of vastly differing microstructures, ranging from blocky to random to perfectly alternating, it cannot provide the level of distinction described above. To achieve this goal, a method that can distinguish "polymeric isomers" is needed. Different levels of distinction can be achieved by various methods. A common method of modeling uses subscripted notation to denote chains with particular features. For example, $P_{n,m,b}$ might represent the number of polymer chains with $n$ units of monomer 1, $m$ units of monomer 2 and $b$ branches. While such a method is ideal for population balance modeling and the application of statistical moments to the chain distribution, it does not give detail about the precise sequencing of the monomer units or branches in the chain. Alternatively one may graphically illustrate the structure of the chain such as in Figure 1, which shows three "polymer isomers" each consisting of five styrene units and six methyl methacrylate (MMA) units. This method is highly descriptive, but is not suited for mathematical interpretation. An equivalent, but more compact representation can be obtained by using Roman letters to symbolize the monomer units (M for MMA, S for styrene). In this representation, strings of letters such as SSSMMSMSM uniquely denote the precise sequence of the chain. This method is still not mathematically based and consequently not amenable to modeling. The method proposed in this article, termed "Digital Encoding of Polymer Chains", utilizes digits to symbolize the particular features of interest on the chain. The number of distinguishable features in a given system dictates the base of the mathematics to be used and the sequence of digits forms a unique string. Such a string can then be translated into a decimal equivalent that provides a unique and compact representation of the chain and all of its features. Similar approaches are utilized in the analysis of chaos by "symbolic dynamics"[2],[3],[4] and in signal processing by "symbolic analysis"[5],[6],[7],[8]. In these approaches streams of data are digitized according to their location with respect to the data mean (0=below, 1=above), or to a number of pre-specified regions. The statistics of sequence of $n$ consecutive data points are then analyzed to detect whether the signal is behaving in a random or structured fashion.

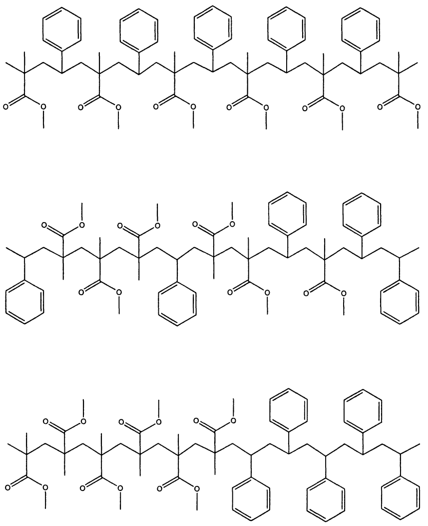


**Figure 1:** Examples of polymer isomers produced in a styrene-methyl methacrylate free radical copolymerization process (initiator fragments at chain ends not shown).

## Digital Encoding Technique

Consider a simple copolymerization scheme involving styrene (S) and acrylic acid (A) monomers (by way of example). The exact sequence of monomers on any polymer chain, such as ASAASSA, can be completely described by the use of two digits (0 and 1 for styrene and acrylic acid, respectively), thus we may use binary mathematics to denote the chain architecture. The binary code representation of our example chain is 1011001. This has a unique decimal equivalent *'d'* that can be expressed by:

$$d = \sum_{i=1}^{n} 2^{i-1} v_i \tag{1}$$

where $n$ is the length of the sequence, $i$ is the position along the sequence (starting from the rightmost bit), and $v_i$ is the digit value (0 or 1) at position $i$.

This technique can be used to graphically represent SLD results obtained from probabilistic calculations. These are based on the "long chain" approximation which states that if the polymer chains are sufficiently long, then the SLD will only be affected by propagation reactions, and thus are a unique function of the reactivity ratios. In this approach, the probability of occurrence of any specific sequence of monomers, say 1011001, is calculated[1] as the combined probability that a propagating radical terminated in 1 will add 0, then 1, then 1, and so on... The statistics thus generated for all types of sequences of a particular length are then studied as these can be compared to measurements by NMR techniques. Application of the proposed "digital encoding technique" allows the numerical distinction of each type of sequence, and provides a suitable method for graphical representation, which was not previously possible. This process is illustrated in Figure 2, for a number of different reactivity ratio combinations and a sequence length=6. Figures 2a-2e show the effect of increasing the reactivity ratios when these are kept equal, while 2f depicts a case of unequal reactivity ratios. It is clear that at low values of $r_1$ and $r_2$, alternating copolymer is obtained as evident from the dominance of the two alternating sequences (010101) and (101010). As the reactivity ratios increase, these are still the most probable but the other types are catching up, until at $r_1 = r_2 = 1$ a totally random copolymer is obtained. Further increase beyond this point induces blockiness to the copolymer as 000000 and 111111 dominate.

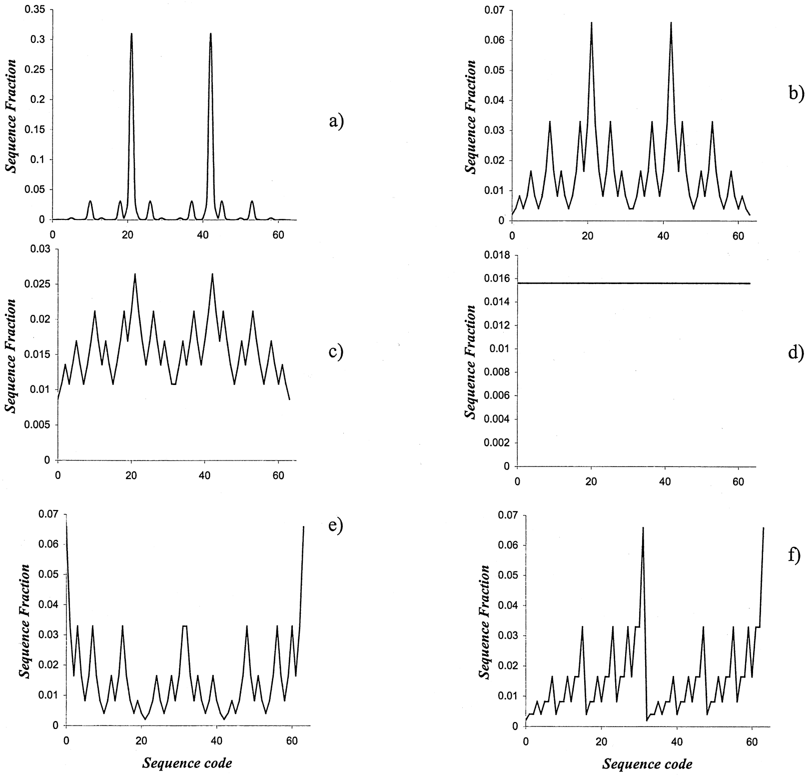


**Figure 2:** SLD for n=6. a)-e): r1=r2=0.1, 0.5, 0.8, 1, 2, respectively. f): r1=0.5, r2=2.

## Application to Population Balance Modeling

Figure 2 illustrates an aspect of usefulness of "digital encoding", but we should not lose sight of the fact that the main objective is to provide distinction at the level of polymer chains, not simply monomer sequences. When attempting to extend the methodology to polymer chains many issues arise and care has to be taken to avoid inconsistencies. These issues include:

- *Accounting for chain end entities*: In most cases these can be assumed to be homogeneously distributed, and thus can be post-assigned to the specific chains. When this assumption is not valid, multiple populations need to be modeled and an increase in the number of model equations will result.
- *Chain orientation convention*: Since radical chains grow by propagation, a direction of growth must be assigned. The convention used is to assume that the chain grows from left to right, i.e. to always place the radical center at the rightmost end of the code, e.g. 1011001*. Dead polymer chains are assumed to inherit the orientation of a terminating radical.
- *Eliminating duplication*: Since some chains can start (on the left side) with a '0', then it is possible to have two chains of differing lengths but of identical codes. This duplicity can be stamped out by always adding a "chain length bit" to the left-hand side of the code, which is equivalent to programming the chain length $n$ into a new, unique code $s$, which can be related to $d$ by $s = d + 2^n$.

Application of these conventions to a simple copolymerization scheme is illustrated in Figure 3, in which the subscripts for each polymeric species denote $n$ and $d$, respectively (note that knowledge of these two uniquely defines $s$). In Figure 3, the change in the chain code upon reaction is indicated for each step. Because of space considerations, the complete modeling equations are discussed elsewhere,[9] but the logic behind the binary operations will be given for each. In the propagation reaction, a new monomer unit is inserted to the right of an existing binary string, which thus has to be fully shifted by one bit to the left. This is mathematically equivalent to multiplying each bit in the string (and consequently the whole number) by a factor of 2, then adding the proper code for the acquired monomer unit (0 or 1). It should be noted that the identity of the radical (and thus the specific propagation reaction it undergoes) is revealed by whether its code is odd or even. Thus an even code indicates a radical center involving monomer of type '0', and an odd code indicates type '1'. It should also be noted that this technique is not limited to the "terminal model" of copolymerization, as the "penultimate unit effect" [10] can be easily incorporated by keeping track of the last two bits of each string.

*Initiation*

$$I \xrightarrow{k_d} 2 \cdot R_{0,0}$$

$$M_1 + R_{0,0} \xrightarrow{k_{i1}} R_{1,0}$$

$$M_2 + R_{0,0} \xrightarrow{k_{i2}} R_{1,1}$$

*Propagation*

$$R_{i,d} + M_1 \xrightarrow{k_{p11}} R_{i+1,2d}$$

$$R_{i,d} + M_2 \xrightarrow{k_{p12}} R_{i+1,2d+1}$$

$$R_{i,d} + M_1 \xrightarrow{k_{p21}} R_{i+1,2d}$$

$$R_{i,d} + M_2 \xrightarrow{k_{p22}} R_{i+1,2d+1}$$

*Termination*

$$R_{i,d} + R_{j,e} \xrightarrow{k_{td}} P_{i,d} + P_{j,e}$$

$$R_{i,d} + R_{j,e} \xrightarrow{k_{tc}} P_{i+j,\,2^j d+\bar{e}}$$

**Figure 3:** Copolymerization Scheme ($\bar{e}$ is the mirror image of e).

Termination reactions involve the formation of dead polymer chains for which the code is determined by the codes of the terminating radicals. Termination by disproportionation only affects the chain end identity, which causes no change in the binary sequence code inherited from the parent radical. Termination by combination on the other hand is a little more complicated, as it involves the production of one dead polymer chain by combining two live radicals. Because of the convention of assigning the radical to the right hand side of the growing chain, this must involve the "flipping" of the binary sequence of one of the radicals to translate the code to its mirror image. This is always applied to the radical considered last, duplication of this operation is prevented however as the summation terms for termination by combination require division by a factor of 2. Additionally, for each binary sequence there are a number of possible combinations of chain fragments that result in the same polymer chain size and binary code. Thus the model equation must consider all possible combinations of sequence codes *d* and *e* that result in the formation of a specific sequence code *g*. This process is illustrated in Figure 4 for a chain with *n*=7 and binary code 1011001.

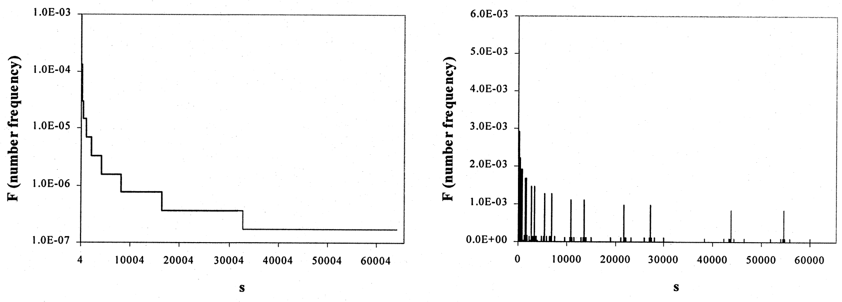
| ***d*** | $\bar{e}$ | ***e*** |
|---|---|---|
| **1** | **011001** | **100110** |
| **10** | **11001** | **10011** |
| **101** | **1001** | **1001** |
| **1011** | **001** | **100** |
| **10110** | **01** | **10** |
| **101100** | **1** | **1** |

**Figure 4:** The six combination possibilities that produce chains with n=7, binary code 1011001, and decimal equivalent g =89.

## Results and Discussion

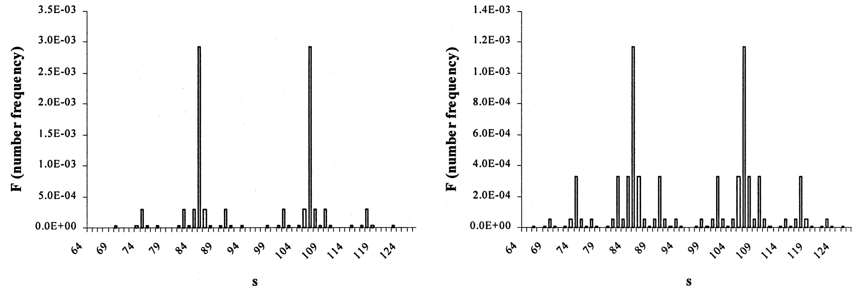
The utility of the modeling technique is demonstrated by the simulation of a process of copolymerization in a CSTR at steady state. The resulting model equations form a set of algebraic equations representing the population balances for each chain length ‘n’ and code ‘d’ combination. The equations are solved sequentially as the reaction mechanism considered does not include any feedback steps. Figure 5 (a,b) shows the results for two cases where termination is solely by disproportionation and where random ($r_1$= $r_2$=1) and a near-alternating ($r_1$= $r_2$=0.1) copolymers are produced, respectively. The frequency is plotted against the modified code “*s*” for a case where $k_{11}$=$k_{22}$, $[M_1]_0$=$[M_2]_0$ and the average degree of polymerization, $DP_n$ = 46. For the random copolymer (Fig. 5a) there is an equal probability of forming all sequences at a particular chain length. The probability of forming this chain length decreases with size on a number basis. For the alternating copolymer (Fig. 5b) only certain sequences are probable and they appear in pairs due to symmetry. For example, for chain size 13 (*s*= 8192 to 16383) the strings 0101010101010 and 1010101010101 (after stripping the “chain length bit”) are most probable, which indicates an alternating structure. These two chains would normally be indistinguishable except for the presence of the initiator fragment on different ends of the chain.

The effect of the termination mechanism is illustrated in Figure 6, which compares termination by disproportionation and combination for the case of an alternating copolymer ($r_1$=$r_2$=0.1). For disproportionation (6a) two codes are most dominant as before. Since disproportionation does not violate the “long chain approximation”, this result agrees with the traditional SLD calculations based on probabilities and reactivity ratios. However, in the case of combination, Fig. 6b indicates an increase in the probability of formation of a number of additional sequences, such as 100101 or 110101. These chains form from the combination of radicals ending with identical monomers. The prediction of these chains is not possible with conventional techniques that only consider the effect of propagation events and not termination on the SLD.



a) b)

**Figure 5:** Sequence code distributions for copolymerization with termination by disproportionation, $k_{11}=k_{22}$; $[M_1]_0=[M_2]_0$; $DP_n = 46$, a) $r_1=r_2=1$ b) $r_1=r_2=0.1$.



a) b)

**Figure 6:** Sequence code distributions for copolymerization with $r_1=r_2=0.1$. a) Termination by disproportionation, b) Termination by combination.

In cases where the reactivity ratios are not equal or where the monomer concentrations differ, more complicated characteristic patterns evolve, these will be discussed in more detail elsewhere[9].

## Potential for Further Applications

The extension of the applicability of the proposed digital encoding technique is generally straightforward, but will eventually run into two major *current* limitations. The first is computational and has to deal with the geometric increase in the number of model equations as longer chain lengths are considered. It is possible however that pattern recognition and data compression techniques can be used to obtain useful information at high chain length without solving the full set of equations, suggestions are made in [9].

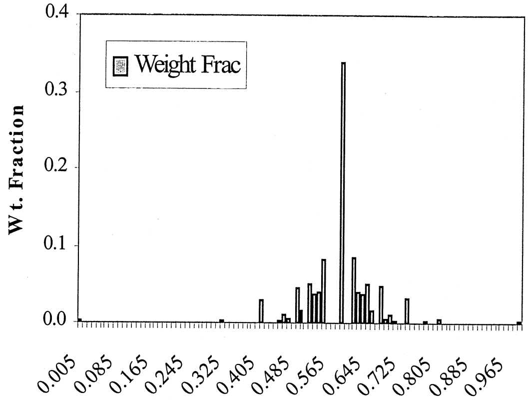


**Figure 7:** Composition distribution for disproportionation with $k_{11}=k_{22}$; $[M_1]_0=[M_2]_0$; $DP_n = 46$, $r_1=0.1$, $r_2=0.1$.

The second limitation is the current lack of suitable analytical techniques for characterization of the 'polymer isomers' that are modeled by digital encoding. Luckily, the digital code formulation of the population balance equations allows convenient extraction of useful information about the chains such as the composition distribution, or the prediction of their MALDI-TOF spectrum (Matrix Assisted Laser Desorption Ionization-Time Of Flight Mass Spectroscopy). Figure 7 presents the composition distribution calculated up to a chain length of 15 in a copolymer sample. Notice the discrete nature of this distribution, resulting from the consideration of the short chain sizes only. Taken over all chain lengths, the composition distribution would normally appear as a continuous function. Figure 8 shows the simulated MALDI-TOF spectra of a random copolymerization system with termination by disproportionation only. The characteristic molecular masses appear at regular intervals indicating a sub-distribution of polymer composition at each specific chain length. The molar masses for the two monomers were taken to be 72 and 104, respectively.

Other analytical techniques can thus be designed to capitalize on the rich level of information found at the low end of the chain length distribution. The possibilities include multi-dimensional chromatography,[11] or the analysis of SLD by what we term "Preparative NMR",[9] a technique that involves the use of NMR techniques on low end fractions of the distribution.

The digital encoding technique is also extendable to polymerization systems involving other reactions. For example, backbiting followed by chain scission can be modeled by extracting the oligomer code from the chain end. Terpolymerization can be modeled by changing the base of the arithmetic math from 2 (binary) to 3 (ternary) and thus using 0, 1, 2 as the bit identifiers. An example terpolymer string could be 1211021, which has a decimal equivalent *d*=1330. In "long chain branching" the technique can be used to model the "primary chain" population, by "chopping" the chain at the branch points and considering these as a comonomer. The encoding process for the primary chains is shown in Figure 9. Also "short chain branching" in polyolefins may be modeled in one of two ways, as illustrated in Figure 10. For example in the case of polyethylene one may use a 0 to denote the polyethylene monomer and a 1 to indicate a short chain branch. Depending on the transfer mechanism (1,3 or 1,5 for example) different types of branches may be formed and can be represented in more detail by shifting to higher base math and uniquely defining the branch structure (butene, hexene, etc.)

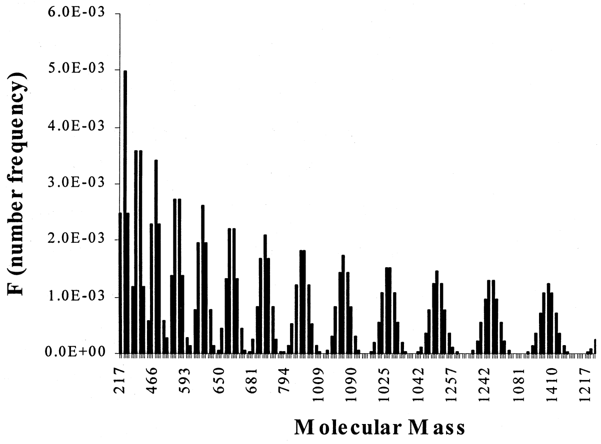


**Figure 8:** MALDI-TOF spectrum for disproportionation with $k_{11}=k_{22}$; $[M_1]_0=[M_2]_0$; $DP_n = 46$, $r_1=1$, $r_2=1$.

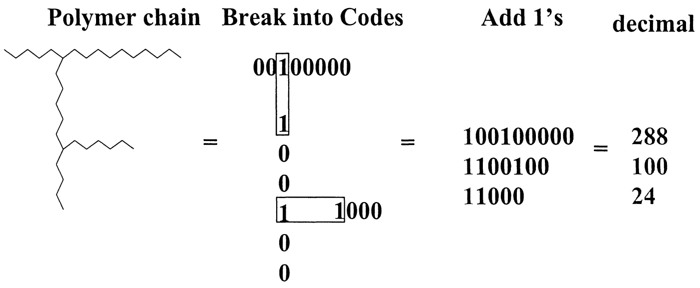


**Figure 9:** Example of Digitizing Long Chain Branching.

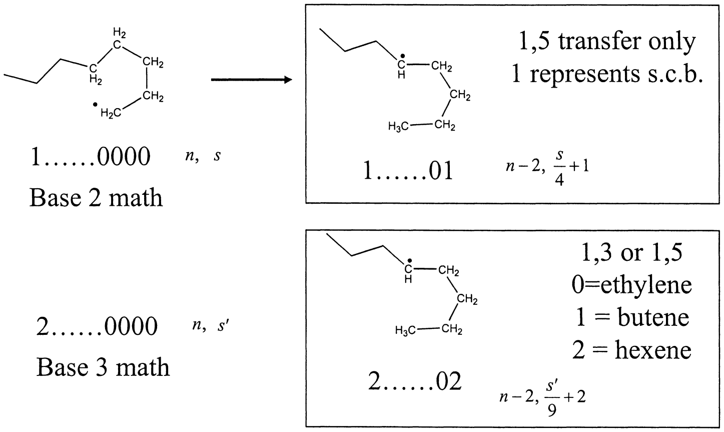


**Figure 10:** Example of a Digitizing Short Chain Branching.

## Conclusions

The digital encoding technique proposed in this article provides a powerful methodology for distinguishing polymer chains at the level of "polymeric isomers". Further application of this method coupled with advances in polymer synthesis and characterization techniques could lead to better designed polymer products.

[1] N. A. Dotson, R. Galvan, R. A. Lawrence, M. Tirrell, *"Polymerization Process Modeling"*, VCH Publishers, New York, **1996**.

[2] J. P. Crutchfield, N. H. Packard, *Physica D*, **1983**, 7, 201.

[3] X. Z. Tang, E. R. Tracy, A. D. Boozer, A deBrauw, R. Brown, *Phys. Rev. E*, **1995**, 51, 3871.

[4] M. Lehrman, A. B. Rechester, R. B. White, *Phys. Rev. Lett.*, **1997**, 78, 54.

[5] C. E. A. Finney, J. B. Green, C. S. Daw, *SAE Paper No. 980624*, **1998**.

[6] C. S. Daw, C. E. A. Finney, K. Nguyen, J. S. Halow, *ASME Congress, Anaheim, CA, USA,* **1998**, p. 221.

[7] C. E. A. Finney, K. Nguyen, C. S. Daw, J. S. Halow, *ASME Congress, Anaheim, CA, USA,* **1998**, p. 405.

[8] C. S. Daw, C. E. A. Finney, M. B. Kennel, *Physical Review E*, **2000**, 62, 1912.

[9] J. A. Debling, F. Teymour, *"Modeling Continuous Linear Copolymerization by Digital Encoding"*, **to be submitted, 2001**.

[10] A. D. Jenkins, K. F. O'driscoll, *J. Polym. Sci., A: Polym. Chem.*, **1995**, 33, 1189.

[11] H. Pasch, *Macromol. Symp.*, **this issue**.